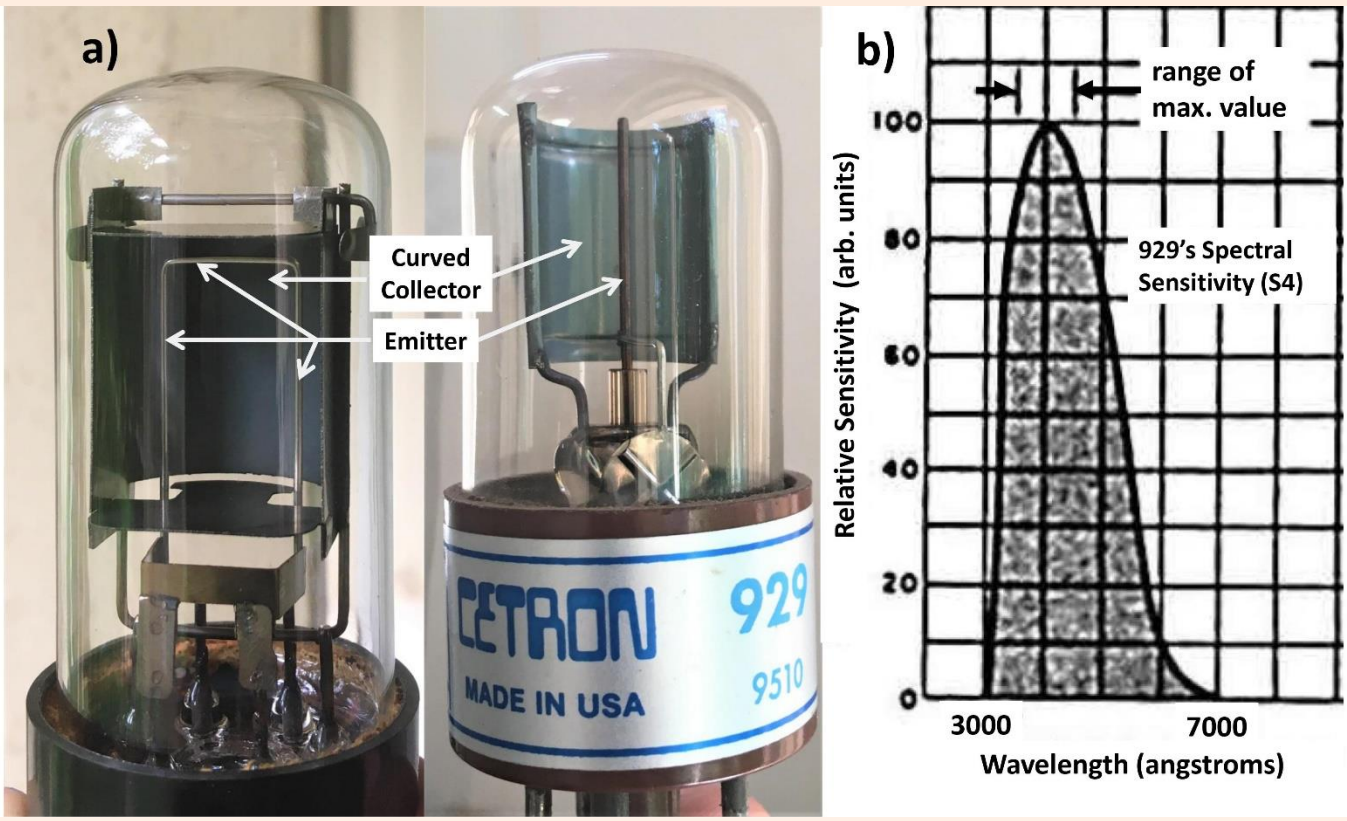
ISSN 0970-5953

Volume 37. No. 4

October - December 2021

PHYSICS EDUCATION



www.physedu.in

Volume 37, Number 4

In this Issue

Understanding Monte Carlo Algorithm and its application to a Bose-Einstein System

Pragati Ashdhir[1], Yash Saxena[2], Adesh Kushwaha[3] and Karun Gadge[4]

[1]Department of Physics, Hindu College, University of Delhi,Delhi, India.
[2]Department of Physics and Astrophysics, University of Delhi, Delhi, India.
[3]Department of Physics, Indian Institute of Technology Gandhinagar, India.
[4]School of Basic Sciences, Indian Institute of Technology Mandi, India.

1.pragatiashdhir@hinducollege.ac.in
2.1980234@pg.du.ac.in
3.adesh.kushwaha@iitgn.ac.in
4.v19090@students.iitmandi.ac.in

Submitted on 10-05-2021

Abstract

In this paper we study the Monte Carlo Algorithm using two different sampling methods, namely, Direct sampling and Markov chain sampling. In the Markov chain sampling method, Metropolis algorithm is used.

The paper explains the Monte Carlo algorithm and the different sampling approaches in a simplified manner. In connection with this we explore the theory of evaluation of integrals and Hamiltonians. To further clarify the tools, a problem of Boson Gas under a harmonic potential trap is chosen and solved using the Monte Carlo Method. The results obtained using different approaches are compared and presented in the form of plots. The aim of this exercise is to efficiently deliver the concept of Monte Carlo method to the graduate and post-graduate physics students and aid in enhancing their computational skills.

1 Introduction and Background

Today, computational methods are an integral part of problem solving in the varied fields of Theoretical Physics Research. One such powerful computational method is the Monte Carlo Algorithm method. The scope of Monte Carlo method [1] is humongous and a complete presentation of the same in a paper like this cannot be justified. Nevertheless, we have attempted to present a simplified usage of the method for students to understand how to implement it and equip

them to develop it further to solve more complicated problems of contemporary research.

At its core, the Monte Carlo method is a statistical approach to computing integrals. It is a numerical technique that uses random numbers and statistical probability to solve problems where there is absence of analytic solutions due to a large number of entities being present (like the system of real gases) or there exist very complex interactions amongst them. This method of random sampling generally provides reliable approximate solutions. When we look back in history, the first article 'The Monte Carlo Method' by Metropolis and Ulam was published in 1949. It is named after the city of Monte Carlo in Monaco, which is famous for all kinds of gambling.

Markov chain Monte Carlo and Metropolis Algorithm[1]: In the year 1906, Andrey Markov first introduced the Markov chain as a stochastic process(containing random variables whose transition depends upon certain assumptions). *The main idea behind the Markov chain is that the next step of the random variable depends upon the current state only, that is, it's a memory-less process where future action depends on the current action.* For general stochastic simulation techniques known as Markov Chain Monte Carlo, the Metropolis [2] is one of the many algorithms for constructing chains.

In Section A, we have introduced the concepts of different samplings to the reader. After that we have explained the use of these sampling methods in computation of integrals.

In section B, we have chosen a physical problem of Boson gas under a 1-Dimensional Harmonic Potential and have used the concept of Monte Carlo sampling both by Direct method and Markov Chain to estimate the energy eigenvalue for the ground state. The computed energy values are then compared with the exact value obtained by analytical methods.

2 Section A

2.1 What is Sampling?

Sampling is a process where the set of entities is collected from a larger set. Each entity has an equal probability of being picked randomly. Next, we need to understand the two fundamentally different sampling approaches: Direct Sampling and Markov Chain Sampling.

2.2 Direct Sampling:

To understand direct sampling [3] we take up the problem of estimating the value of the constant π.

2.2.1 Motivation and Simple Example: Estimating π

Suppose we have a **square** dartboard and a novice is made to randomly throw darts across the board. After a large number of hits, we inscribe a circle inside the dartboard

as shown in Fig. (1) and calculate the ratio between the number of hits inside the circle to that inside the square.

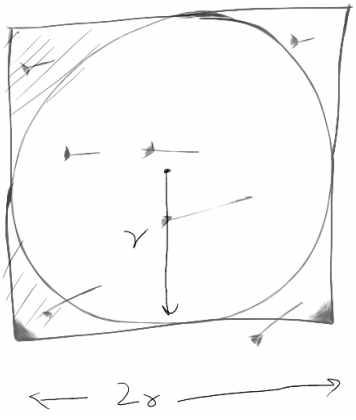


Figure 1: Darts Game for π estimation

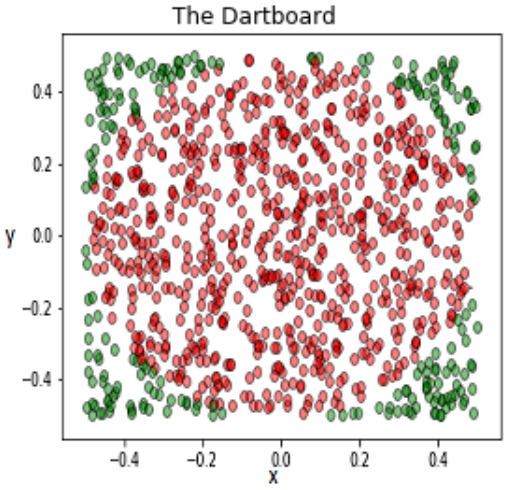


Figure 2: Darts Game Simulation

Using a probabilistic approach, the given ratio can be used to estimate the value of π as follows:

$$f = \frac{N_{circle}}{N_{square}} = \frac{Area_{Circle}}{Area_{Square}} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} \quad (1)$$

We can simulate the dart game using a random number generator code on our computer and get the value of pi. Fig. (2) shows the random points in one such simulation. Similarly, the value of other constants [4] can be estimated.

As depicted in Fig.(3), when we have fewer numbers of points, our estimates vary much more wildly and are much deviated from 3.1415926. However, the guesses from our different runs are all seen to be equally distributed around the correct value, thereby, minimising the chances of any systematic error in our estimate.

In the given simulation, the points required are randomly generated within a defined region. *This approach is the Monte Carlo Direct Sampling method, wherein, random values are generated without any history of the last value/values.* New values are generated independently and are accepted within an allowed area/region.

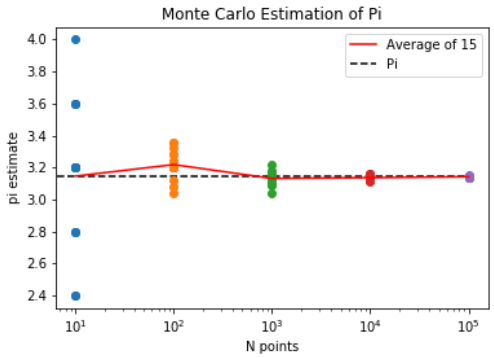


Figure 3: Monte Carlo Estimate of π

2.3 Markov chain Monte Carlo and Metropolis Algorithm:

Markov Chain Monte Carlo is a sampling method in which the generation of a new random number is governed by a previously generated random value. The parameters responsible for the new value are randomly generated, which are then added to or subtracted from the last value. This sampling method can also be understood by reconsidering the example of estimating the value of π by the dart game **with a twist**.

Imagine the dart board to be a very big playground and a person standing on it is throwing darts as javelin in random directions as depicted in Fig.(4) through Fig.(8).

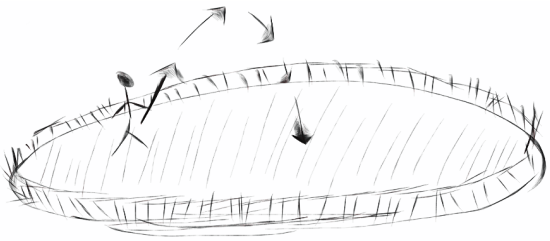


Figure 4: Dart Game with a twist

He first stands at a randomly chosen position on the playground and throws a javelin in a completely random direction. Next, he walks to the location where the javelin fell and throws another javelin in a completely random direction from this new position.If the javelin falls out of the region boundary, he continues to stand on the current position and throws a javelin again in completely random direction.

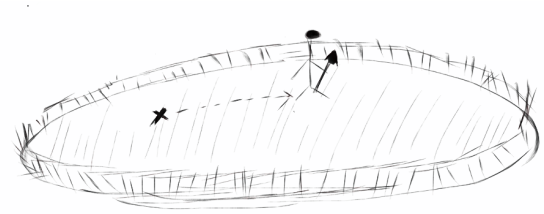


Figure 5: Throw the javelin and occupy the new position for the next throw

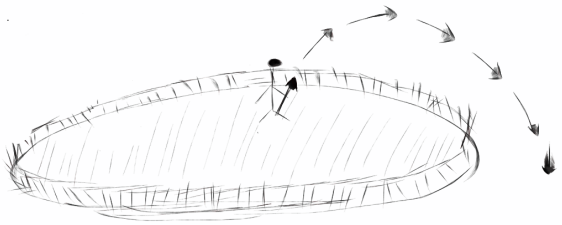


Figure 6: If the javelin is outside the ground, re-throw the javelin from the same position.

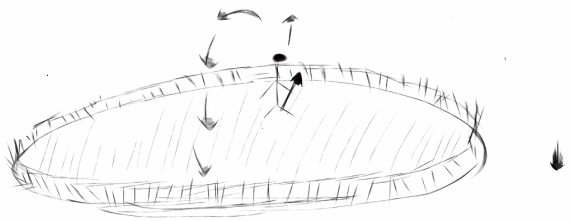


Figure 7: If the javelin is inside the ground, occupy the new position for the next throw.

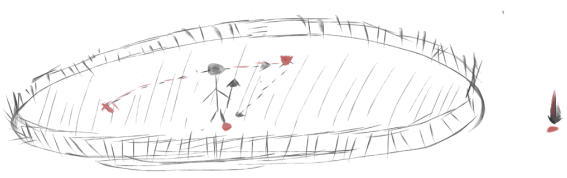


Figure 8: Sampling of the board/ground is shown as red dots

Using the described method, we sample the entire board/ground by generating a random walk [1] of points $(x_1, x_2, ...x_n)$ and hence estimate the value of π.

The Markov Chain sample asymptotically converges to the desired probability distribution, say $\pi(x)$, whereas, in direct sampling there is a direct access to $\pi(x)$. Consider the case of a random walker whose position is x_i, where i is the step of the Markov Chain Sampling. Let Δ be the randomly chosen difference between x_i & x_{i+1}. The acceptance of Δ to generate x_{i+1} is determined by the value of a parameter ω, defined as

$$w = \frac{\pi(x_{i+1})}{\pi(x_i)} \tag{2}$$

If $w \geq 1$, we accept the change and let $x_{i+1} = x_i + \Delta$, otherwise, $x_{i+1} = x_i$ and the above process is repeated.

2.3.1 Evaluation of Integrals:

The Monte Carlo method can also be used to compute integrals as explained below [5]. Let the integral to be computed be:

$$\int_0^1 h(x)dx \tag{3}$$

A large number of sample points say, $x_1, x_2, x_3..., x_n$ are chosen randomly from a uniform distribution in the range (0,1). The function $h(x)$ is computed at these points to yield a set of values $h(x_1), h(x_2)...h(x_n)$. The mean $\langle h(x) \rangle$ of these values is the computed integral value.

This method works since uniform random variables have a special property that their integral is unity over the uniform range of bound i.e. (0,1).

$$\int_0^1 f(x)dx = 1 \tag{4}$$

Now, for any function $h(x)$ if $f(x)$ is the uniform distribution, then using equation (4) we get,

$$\langle h(x) \rangle = \int_0^1 h(x)f(x)dx = \int_0^1 h(x)dx \tag{5}$$

or,

$$\langle h(x) \rangle = \frac{1}{N} \sum_{i=0}^{N-1} h(x_i) \tag{6}$$

where, N represents the sample size.

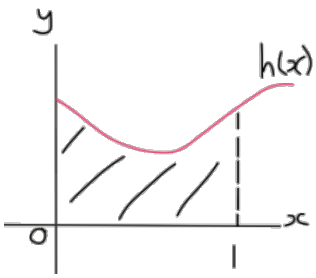


Figure 9: An example of function h(x).

Hence, we are able to compute definite integral by finding a sample mean.

2.3.2 Generalization of Evaluation of Integral for Arbitrary PDF

The formula given in equation (6) works only if the PDF (Probability Distribution function) of the random variable X is uniform. We can extend Monte Carlo integration to random variables with arbitrary PDFs. The more generic formula is given by

$$\langle h \rangle = \frac{1}{N} \sum_{i=0}^{N-1} \frac{h(X_i)}{pdf(X_i)} \tag{7}$$

The generalized expression in equation (7) can be used to solve Quantum Mechanical systems. Before diving into the actual problem solving, we need to understand the concept of Quantum Monte Carlo.

3 Quantum Monte Carlo

Most quantum mechanical problems of interest consist of a large number of interacting electrons and ions or nucleons. The total number of particles (N) is usually large and an exact solution cannot be found easily. In quantum mechanics, we can express the expectation value of a given operator $\langle \hat{O} \rangle$ for a system of N particles by equation (8), where Ψ is the wave function describing a many-body system.

For a multidimensional integral as in equation (8), the Monte Carlo methods are ideal for obtaining expectation values of quantum mechanical operators [6]. Now, our problem is that we do not know the exact wave function Ψ. So, we would invoke the concept of quantum mechanical Variational Principle in to generate a trial wave function and then attempt to perform a variational calculation of the various observables, using Monte Carlo methods.

$$\langle \hat{O} \rangle = \frac{\int dr_1 dr_2 ... dr_N \Psi^*(r_1, r_2, ..., r_N) \hat{O}(r_1, r_2, ..., r_N) \Psi(r_1, r_2, ..., r_N)}{\int dr_1 dr_2 ... dr_N \Psi^*(r_1, r_2, ...r_N) \Psi(r_1, r_2, ..., r_N)} \tag{8}$$

where $r_1, r_2, ..r_N$ are the generalized coordinates of Ψ

3.1 Variational Method in Quantum Mechanics:

According to the Variational Method, given any normalized function Ψ (that satisfies the appropriate boundary conditions), the expectation value of the Hamiltonian represents an upper bound to the exact ground state energy.[7]

$$E_o \leq \langle \hat{H} \rangle, \tag{9}$$

where, E_o is the ground state energy.

For a brief illustration of the method, let us find the ground state energy of one-dimensional harmonic oscillator, whose Hamiltonian is given by:

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2 \tag{10}$$

where, ω, m are the angular frequency and mass of the oscillator respectively. As a trial wave function, we pick a function that satisfies the boundary condition of the exact ground state wave function. The boundary condition is that the wave function vanishes at positive and negative infinities. Hence, we choose the trial wave function as,

$$\Psi = Ae^{-\alpha x^2} \tag{11}$$

where, α is the variational parameter and A is the normalization constant.

We now introduce,

$$\langle \hat{H} \rangle = \langle \hat{T} \rangle + \langle \hat{V} \rangle \tag{12}$$

where, $\hat{T}$ and $\hat{V}$ are the Kinetic Energy and Potential energy operators respectively, defined as,

$$\hat{T} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} \tag{13}$$

giving the expectation value of kinetic energy as,

$$\langle \hat{T} \rangle = \int \Psi^* \hat{T} \Psi dx = \frac{\hbar^2 \alpha}{2m} \tag{14}$$

and

$$\hat{V} = +\frac{1}{2} m\omega^2 x^2 \tag{15}$$

giving

$$\langle \hat{V} \rangle = \int \Psi^* \hat{V} \Psi dx = \frac{m\omega^2}{8\alpha} \tag{16}$$

as the expectation value of potential energy. Minimizing the Hamiltonian with respect to the parameter alpha(α), we get,

$$\frac{d}{d\alpha}\langle \hat{H} \rangle = 0 \implies \alpha = \frac{m\omega}{2\hbar} \tag{17}$$

Using the above value of α, we get,

$$\langle \hat{H} \rangle_{min} = \frac{1}{2}\hbar\omega \tag{18}$$

as the exact ground state energy.

Having briefly discussed the Variational Principle, we now look back at Quantum Monte Carlo using the Variational Principle.

3.2 Quantum Variational Monte Carlo:

Given a Hamiltonian H and a trial wave function $\Psi(r)$, the variational principle states that the expectation value of $\langle H \rangle$ is

$$\langle \hat{H} \rangle = \frac{\int dr \Psi^*(r) \hat{H}(r) \Psi(r)}{\int dr \Psi^*(r) \Psi(r)} \tag{19}$$

Thereafter, we vary α (variational parameter) according to some minimization algorithm to find the best upper bound to the ground-state energy.

Now, having all the computational tools required to solve a quantum system, we proceed in Section B to solve a Bosonic system in a 1-D Harmonic trap.

4 Section B

4.1 Problem Statement: Bosons in 1-D Harmonic Trap

A class of particles which follows the Bose-Einstein statistics are classified as Bosons. This concept was first proposed by Satyendra Nath Bose, as an attempt to tackle the limitation of the Maxwell-Boltzmann distribution for the microscopic particles at all scales. Along with Albert Einstein, the duo developed a new statistics, now known as the Bose-Einstein statistics. This statistics governs the behaviour of non-interacting, indistinguishable particles with integral value of spins. *Such particles are known to have symmetrical wave functions.*

In this paper, we have chosen a Boson system with a large number of particles under the influence of a 1-dimensional har-

monic potential. In other words, we consider a Boson gas with N particles trapped inside a 1-Dimensional Harmonic potential well. The problem is to determine the energy eigen values of the trapped gas using the Monte Carlo Direct sampling approach and Markov-Chain sampling method. The Monte Carlo method results are then compared with the analytically obtained value.

Even though the given problem of a bosonic system is a standard one with a closed form solution, the concept of Monte Carlo sampling can be very well explained with it. There are various systems wherein the trapping potential [8] is much more complicated and the Monte Carlo method works very well for those systems.

4.2 The System, Wavefunction, Potential and Hamiltonian:

We consider N Bosons with mass m, under the influence of 1-Dimensional Harmonic Potential. Since the system is non-interacting, the internal potential is not taken into account.

The 1-Dimensional Harmonic potential is given by,

$$V(x) = \frac{1}{2}m\omega^2 x^2 \tag{20}$$

where, $\omega = \sqrt{\frac{k}{m}}$ and k is the force constant characterizing the given harmonic potential. The analytically obtained wave function of the particles under the influence of such a potential is given by [7]:

$$\Psi_n(x) = (\frac{m\omega}{\pi\hbar})^{1/4}\frac{1}{\sqrt{2^n n!}}H_n(x)e^{-x^2/2} \tag{21}$$

where the integer n labels the different energy states. The functions $H_n(x)$ are known as Hermite Polynomials. The ground state (n=0) wave function is given by,

$$\Psi_0(x) = (\frac{m\omega}{\pi\hbar})^{1/4}e^{-x^2/2} \tag{22}$$

and is depicted in Fig. (10).

Wavefunction.png

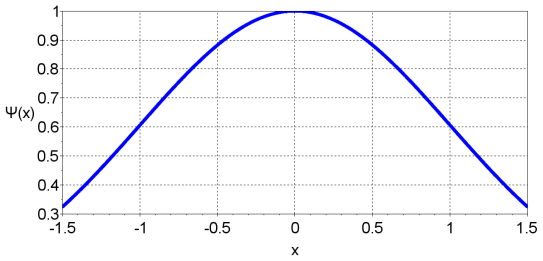


Figure 10: Harmonic Wavefunction for ground state (n=0)

Alternatively, a trail wave function can also be obtained using the concept of Variational Monte Carlo Method (Section 3.2). Accordingly, the expectation value of the Hamiltonian [6] is given by:

$$\langle\hat{H}\rangle = \frac{\int dr_1...dr_N\psi_n^*\psi_n(\frac{\hat{H}\psi_n}{\psi_n})}{\int dr_1...dr_N\psi_n^*\psi_n} \tag{23}$$

The general form of equation (23) has already been discussed in the section 2.3.2. Since we are working with discrete system of random points, the finite energy is calculated in the local region of each random point selected. So, we need to convert equation (23) to discrete form as follows:
A set of M particle configurations, $\pi_1, \pi_2, ...\pi_M$ is considered which corresponds to probability distribution $|\Psi_n|^2$.

The equivalent discrete equation can be shown to be,

$$\langle \hat{H} \rangle > \sim \frac{1}{M} \sum_{n=1}^{M} \left(\frac{\hat{H}(\pi_n)\Psi_n(\pi_n)}{\Psi_n(\pi_n)} \right) \quad (24)$$

Next, the above equation is solved by choosing some random sample using the Monte Carlo direct sampling technique and also by using the Metropolis Random Walk algorithm.

5 Evaluation of Ground State Energy Eigenvalue

5.1 Analytical Treatment

The energy eigenvalues of the particles under the influence of harmonic potential is given by,

$$E_n = (n + 1/2)\hbar\omega \quad (25)$$

where, n=0,1,2..... . Taking $\hbar = 1$ and $\omega = 1$, the ground state energy of the particles is determined to be, $E_0 = 0.5$ eV.

5.2 Monte Carlo Direct Sampling method

Now, we will apply the Monte Carlo method to compute the energy of particles inside the 1-D harmonic trap. In the equation (24), the Hamiltonian of our system is,

$$\hat{H}(x) = \frac{-\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{m\omega^2 x^2}{2} \quad (26)$$

Taking $\hbar = 1, m = 1$ and $\omega = 1$ we have,

$$\hat{H}(x) = \frac{-d^2}{dx^2} + \frac{x^2}{2} \quad (27)$$

In order to calculate the expectation value of $\hat{H}$ given by equation (27), we need to consider a function Y, defined as,

$$Y(x) = \frac{\hat{H}(x)\Psi(x)}{\Psi(x)} \quad (28)$$

Using $\Psi(x)$ from equation (22), we get

$$\hat{H}(x)\Psi(x) = \frac{-d^2 e^{\frac{-x^2}{2}}}{dx^2} + \frac{x^2}{2} e^{\frac{-x^2}{2}} \quad (29)$$

Using equation (28) and equation (29) we get,

$$Y(x) = 1 - \frac{x^2}{2} \quad (30)$$

The Maxima of Y(x) is given by:

$$\frac{dY(x)}{dx} = 0 \quad (31)$$

Equation (31) gives the maximum value as Y(0)=1. This is the upper limit of the energy curve. To select such a sample we first follow the Monte Carlo direct sampling method discussed in section 2.2.

The sample chosen is in the form of x_i and y_i where i runs from 1 to N. Here, x_i represents the randomly chosen position of the i^{th} particle within some arbitrary range and y_i is the corresponding energy value bounded by the upper limit of Y(x). Out of the random sample of x_i and y_i, only certain points are chosen according to the acceptability method described below:

Whenever a random value of y is generated, it is compared with the corresponding value of H(x), if $y < H(x)$, then x_i is accepted, which in turn means that the corresponding value of Hamiltonian is selected, else it is rejected.

Let the number of accepted values be p, where $p \leq N$. Using these accepted values we estimate the ground state energy of the Hamiltonian by calculating the mean of all the individual Hamiltonian values at each point. Figure (11) depicts the variation of the Hamiltonian with the randomly generated x range using the Direct Sampling Method.

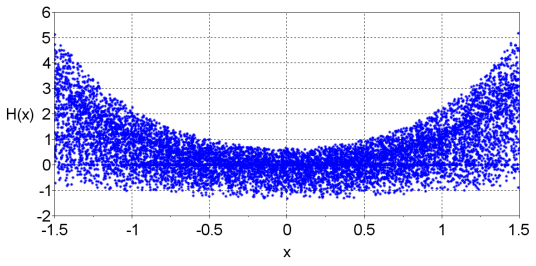


Figure 11: Variation of the Hamiltonian with the randomly generated x range using the Direct Sampling Method

The estimated ground state energy value by Direct Sampling Method (using Scilab [9]) is 0.5415408 eV.

5.3 Markov Chain Monte Carlo method : Random Walk Sampling

In Markov Chain sampling, we have used the ideas discussed in Section 2.3. Let the variable, n_{trials}, represent the number of random walks we use in the program. Initially, we choose a random value of x and y, which are subsequently updated in accordance to Markov Chain Method to generate a random path comprising of p number of accepted points. The process is repeated to generate such multiple random walks ($= n_{trials}$) to cover the entire under the given energy curve. At each of the accepted points (x, y), the value of Hamiltonian is computed and summed to give the total Hamiltonian, which is then divided by $p * n_{trials}$ to give the value of energy of Bosons. The condition of acceptability of a Hamiltonian value for a certain point is the same as discussed in the direct sampling method (Sec 5.2).

5.4 Random Walk Plots

Fig. (12) through Fig. (19) depict the plots of Y(x) and the corresponding Hamiltonians H(x) for an increasing number of random walks (n_{trials} = 1, 2, 10 & 50). The computed values of ground state energy in each case are given in table (1).

Number of Random Walks	Estimated Ground State Energy (eV)	Error(%)
1	0.1718306	65.63
2	0.902497	80.49
10	0.5768063	15.36
50	0.5690777	13.82

Table 1: Computed Values of Ground State Energy for Different Random Walks

It can be seen from the table and the random walk plots that as the number of random walks is increased, i.e., a larger portion of the area under the Y(x) energy curve is

covered, the estimated ground state energy approaches the exact value.

1. For 1 random walk

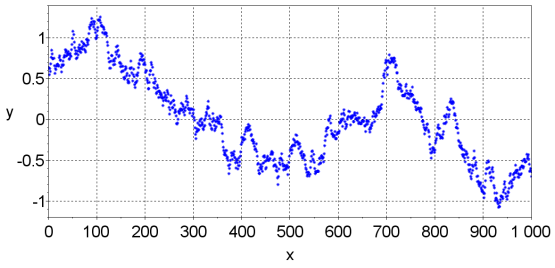


Figure 12: Random Walk-1

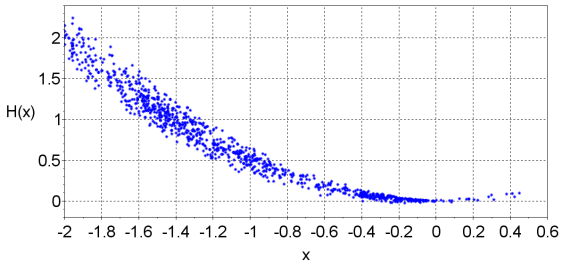


Figure 13: Hamiltonian Plot for Random Walk-1

2. For 2 random walks

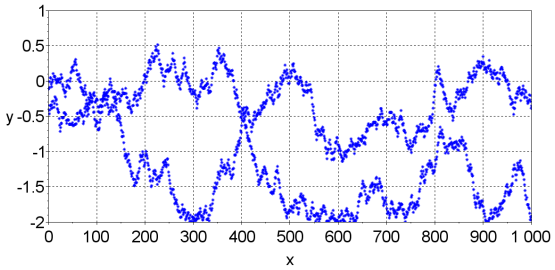


Figure 14: Random Walk-2

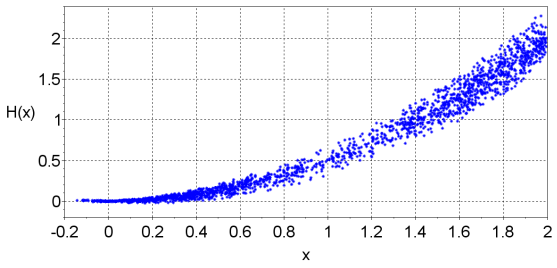


Figure 15: Hamiltonian Plot for Random Walk-2

3. For 10 random walks

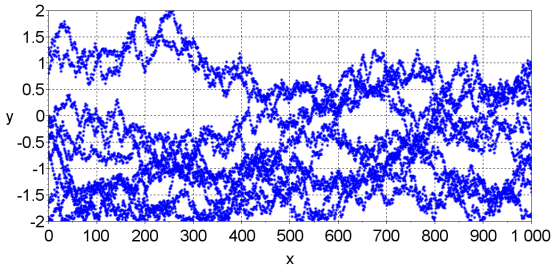


Figure 16: Random Walk-10

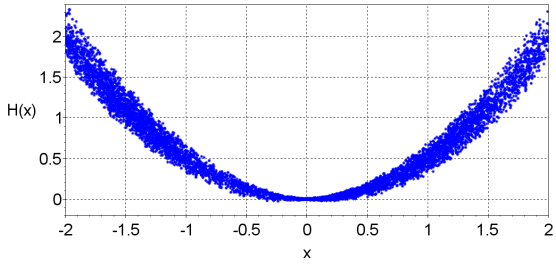


Figure 17: Hamiltonian Plot for Random Walk-10

4. For 50 random walks

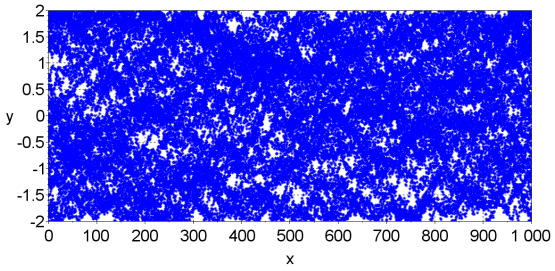


Figure 18: Random Walk-50

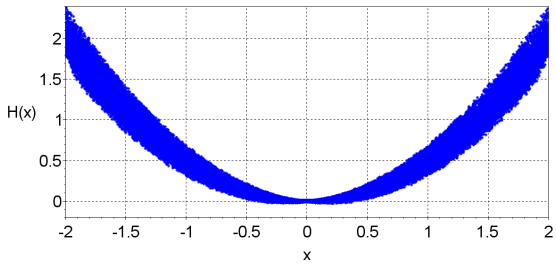


Figure 19: Hamiltonian Plot for Random Walk-50

6 Results and Discussions:

In this paper we have used the two approaches of sampling as described in Section A. We saw through this paper that Monte Carlo method gives us a strong and powerful tool, to solve the Hamiltonians to compute energy eigenvalues. It is quite an elegant concept for the potentials which are complicated and are difficult to solve analytically. Monte Carlo equips us to solve such complicated potentials without getting into tedious mathematics, where chances of er-

rors are more.

The computation of ground state energy of Bosonic gas under the influence of 1-D Harmonic potential using the Monte Carlo direct sampling method and the Metropolis method yielded results in good agreement with the exact value of 0.5 eV.

References

[1] Werner Krauth. Statistical mechanics: algorithms and computations. Oxford University Press, 2012.

[2] David P. Landau. The metropolis monte carlo method in statistical physics. In AIP Conference Proceedings. AIP, 2003.

[3] D. E. Raeside. An introduction to monte carlo methods. American Journal of Physics,42(1): 2026, January 1974.

[4] Pirooz Mohazzabi. Monte carlo estimations of e. American Journal of Physics, 66(2): 138140, February 1998.

[5] S Iskandar. Modified monte carlo method for integral. Journal of Physics: Conference Series, 1462 : 012061, February 2020.

[6] J. L. DuBois and H. R. Glyde. Bose-einstein condensation in trapped bosons: A variational monte carlo analysis. Physical Review A, 63(2), January 2001.

[7] David Griffiths. Introduction of Quantum Mechanics. Prentice Hall, Inc., 1995.

[8] Salvio Jacob Bereta, Lucas Madeira, Vanderlei S. Bagnato, and Monica A. Caracanhas. Boseeinstein condensation in spherically symmetric traps. American Journal of Physics,87(11): 924934, November 2019.

[9]]S. Enterprises, Home - Scilab, Scilab.org, 2017. [Online]. Available: https://www.scilab.org/

New and simple method for the determination of critical damping resistance of moving coil galvanometer

Bhakta Kunwar

Post –Graduate Department of Physics, Sikkim Government College,
Tadong, Gangtok, Sikkim, India-737102.

E-mail: bhaktakunwar@yahoo.co.in

Submitted on 25-02-2021

Abstract

Moving coil galvanometer is one of the most frequently used apparatus in electromagnetic measurements of any undergraduate physics laboratory. The critical damping resistance of a moving coil galvanometer is a crucial parameter for determining its operation either as a dead beat galvanometer or as a ballistic galvanometer. The value of critical damping resistance is provided in the data sheet by some of the standard manufacturers of galvanometer. Ironically, many of the manufacturers do not provide it. Methods of measuring critical damping resistance in a laboratory have been discussed by few authors. However, these methods are meticulous, involved and time consuming from the point of view of a student working within a limited time in a laboratory. In this paper we present a simple and alternate method of determining critical damping resistance of a moving coil galvanometer. Firstly, we present the theory concerned and derive the required working formula. No approximations have been made in the underlying theory while deriving the working formulae. Then, we illustrate the method by performing an experiment. The method is simple and less time consuming.

1. Introduction

A moving coil galvanometer (MCG) is used in a variety of electromagnetic measurements in an undergraduate physics laboratory. It not only offers scope to understand basic interactions in electricity and magnetism but also finds applications in various sensitive measurements. As is well known, an MCG can be used either as dead beat galvanometer or as a ballistic galvanometer (BG). A dead beat galvanometer can measure steady current while a BG can measure a momentary flow of charge.

An MCG can be made to operate in BG mode simply by connecting an external resistance, having a value greater than the so called 'critical damping resistance' (CDR), in series with the galvanometer. An MCG function as a dead beat galvanometer if the external resistance is less than the CDR. The theory of MCG, its construction and design were well known by the beginning of twentieth century [1,2,3,4]. Some of the MCGs are designed to operate in critically damped conditions, the theory and method of designing such MCGs can be found in the literatures, for example in [5]. The value of CDR of a given MCG depends on intrinsic constants of the materials used in its construction such as moment of inertia

of suspended coil, restoring constant of suspension, magnetic field, and area and number of loops in the coil. Methods of measuring CDR of a general MCG in a laboratory have been discussed by several authors [6,7]. These methods are meticulous, involved and time consuming from the point of view of an undergraduate student working within a limited time in a laboratory. We present a simple method of finding CDR in a laboratory by using simple formulae which have been derived without involving too many approximations. The method and the formulae have been tested and verified through experimental investigations.

In the following sections, we briefly describe construction and theory of MCG. Subsequently, we derive a working formula for finding CDR and we describe the procedure for performing the proposed experiment. Finally we illustrate our proposition with experimental data.

2. Construction and theory of the galvanometer

An MCG consists of a movable coil suspended in the magnetic field of permanent magnet. When an electric current is passed through the coil, the latter experiences a torque due to the Lorentz force. The torque rotates the coil until the driving torque is balanced by the restoring torque of the coil suspension. Let N be the number of turns in the coil, A be the area of coil and B be the strength of magnetic field. Let us introduce a constant, $G = NAB$.

If I is the moment of inertia of suspended coil about its axis of rotation, a is the damping constant due to air etc., c is restoring constant, θ is the angle of deflection of the coil after time t of applying a dc emf E to the galvanometer circuit as shown in figure 1, then the equation of motion of the moving system is

$$I\frac{d^2\theta}{dt^2} + a\frac{d\theta}{dt} + c\theta = \frac{G}{R}\left(E - G\frac{d\theta}{dt}\right), \qquad (1)$$

where R, the total resistance of the circuit, is given by $R = R_g + R_e$, R_g and R_e being the resistances of MCG and the external resistance respectively. The equation of motion of the coil can be written as

$$\frac{d^2\theta}{dt^2} + 2p\frac{d\theta}{dt} + q^2\theta = f, \qquad (2)$$

where $$p = \left(\frac{a}{2I} + \frac{G^2}{2IR}\right), \qquad (3)$$

$$q = \sqrt{\frac{c}{I}}, \qquad (4)$$

and $$f = \frac{GE}{IR}. \qquad (5)$$

The behavior of MCG will be overdamped, critically damped or underdamped according as $p^2 > q^2$, $p^2 = q^2$ or $p^2 < q^2$ respectively. In the underdamped condition, that is when $p^2 < q^2$, the motion of the coil is oscillatory and its amplitude is given by

$$\theta = \theta_0\{1 - Ae^{-pt}sin(\omega t + \alpha)\}, \qquad (6)$$

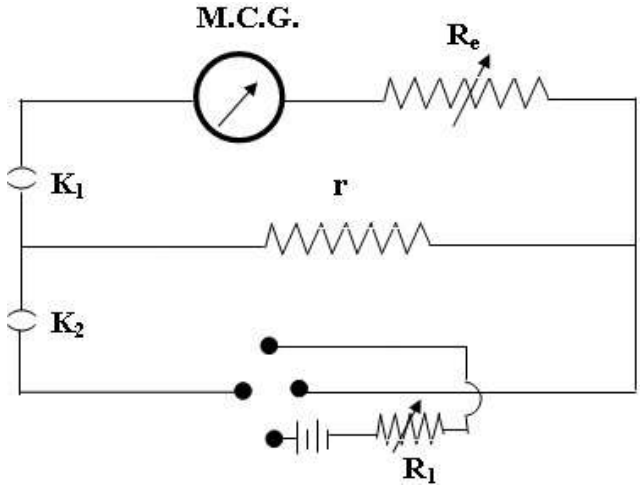


Figure 1. Circuit diagram for the determination of critical damping resistance of an MCG. Low resistance r is used for providing a potential difference across MCG via external resistance R_e. Variable resistance R_1 in the battery circuit is

used for adjusting the emf so as to produce ballistic throw within the scale.

where ω is the angular frequency and is given by $\omega = \sqrt{q^2 - p^2}$. The constants A and α in eq.(6) are given by $A = \sqrt{(p^2/\omega^2) + 1}$ and $\alpha = tan^{-1}(\omega/p)$.

The time period of oscillation $T = 2\pi/\omega$ can be expressed as

$$T = \frac{2\pi}{\sqrt{\frac{c}{I} - \frac{(a+G^2/R)^2}{4I^2}}}. \qquad (7)$$

The eq.(6) represents a damped oscillation about the final steady deflection θ_0.

2.1 About Damped Oscillation

If an MCG is made to operate in underdamped condition and a current is passed through it by inserting both the keys K_1 and K_2 in their positions as shown in figure 1, the coil would undergo oscillatory motion about a mean position. The coil would finally reach a steady state with steady deflection θ_0 .The amplitudes of moving system under this condition are given by eq.(6).
Now, if K_2 is suddenly removed from the circuit keeping K_1 fixed in its position, the coil would undergo damped oscillation about its equilibrium position (zero position of scale) and the variation of θ with time t is then given by

$$\theta = \theta_0 e^{-pt} sin\omega t. \qquad (8)$$

If θ_1 be the first peak throw on one side of the zero, θ_2 the second throw on the other side, θ_3 the third throw (on the side of θ_1), θ_4 the fourth peak throw (on the side of θ_2) and so on. Then,

$$\frac{\theta_1}{\theta_2} = \frac{\theta_2}{\theta_3} = \frac{\theta_3}{\theta_4} = \cdots = e^{pT/2} = x\,.$$

The quantity $\ln x$ is called logarithmic decrement and is denoted by λ as

$$\lambda = \ln x = \frac{pT}{2}\,. \qquad (9)$$

In order to eliminate the individual errors introduced by inaccurate readings due to shifts of the zero of the scale, log decrement λ can be obtained from the relation

$$\lambda = \frac{1}{2n} ln \frac{\beta_1}{\beta_{2n+1}} \qquad (10)$$

where $\beta_1 = \theta_1 + \theta_2$, $\beta_2 = \theta_2 + \theta_3$ and so on. Since $\beta's$ are the sum of the successive deflections on the two sides of the zero of the scale, the errors affecting the individual deflections due to the inaccurate reading of zero are eliminated. See textbook such as Feweks and Yarwood [8].

3. Our proposition: Working Formula and experimental procedure

For a given value of external resistance R_e, one can achieve steady state with both the keys K_1 and K_2 fixed in their positions. Then, if K_2 is suddenly removed from the circuit keeping K_1 fixed, the coil would undergo damped oscillation about the zero position as discussed above. Then one can measure and record the values of θ_1, θ_2, θ_3, etc. and from these measured values of deflections one can find logarithmic decrement λ using eq.(10). One should also measure time period T at this value of R_e. After measuring λ and T for this value of external resistance R_e, we can find the damping term p from eq.(9). Next, we change the values of R_e and repeat the above procedure so as to obtain λ , T and p at different values of R_e. We also find values of λ, T and p at $R_e = \infty$; this is done by taking out both the keys K_1 and K_2 from their positions after achieving steady state. Such oscillations are usually called the open circuit oscillations. If we plot a graph with reciprocal of R along x-axis and p along y-axis, then, we should get a straight line according to eq.(3). The slope of this straight line should be equal to $G^2/2I$ and the intercept on y-axis should give the value of $a/2I$.

Moreover, in the open circuit oscillations, $R_e \rightarrow \infty$ and therefore, $R \rightarrow \infty$. This implies from eq.(3) that the experimentally obtained value of p in the open circuit condition, say p_∞, must also be equal to $a/2I$. So, it is expected that the intercept on y-axis should match with measured value of p_∞.

Now, one can derive from eqs(3) and (7)

$$\frac{c}{I} = p^2 + \frac{4\pi^2}{T^2} \qquad (11)$$

Using the the values of T and p measured at different values of R_e, one can find the values of c/I at corresponding values of R_e from eq.(11). Consequently, an average value, say $(c/I)_{av}$, can be obtained from the above set of values of c/I calculated at different R_e. But, as c and I both are intrinsic constants for a given MCG, the value of c/I and hence the right hand side of eq.(11) should also be a constant. It is worth mentioning here that values of T and p measured in open circuit condition have the lowest error and therefore the value of c/I calculated with these values of T and p is expected to be close to the average value $(c/I)_{av}$.

Now, the condition for critical damping is $p^2 = q^2$. Let this condition be achieved at $R_e = R_{cdr}$; this R_{cdr} is termed as critical damping resistance (CDR). The condition $p^2 = q^2$ can also be expressed as

$$\left\{\frac{a}{2I} + \frac{G^2}{2I(R_g + R_{cdr})}\right\}^2 = \left(\frac{c}{I}\right)_{av}, \qquad (12)$$

where we have used average value of c/I. From eq.(12), one can easily find an expression for critical damping resistance R_{cdr} as

$$R_{cdr} = \frac{G^2/_{2I}}{\sqrt{\left(\frac{c}{I}\right)_{av}} - \frac{a}{2I}} - R_g \qquad (13)$$

The values of terms $G^2/2I$ and $a/2I$ on the right hand side of eq.(13) are obtained from the slope and intercept of the straight line drawn with $1/R$ along x-axis and p along y-axis, while $(c/I)_{av}$ is calculated from eq.(11) as discussed above. Thus, knowing values of T and p at different values of R_e, one can find the value of critical damping resistance R_{cdr} of a given MCG under a given operational condition.

4. Experimental data and results

We present here an experimental data which was carried out hurriedly for the verification of the proposed theory as a first time check. As such, the experimental data and the results eare mainly aimed for illustrative purpose. We took an MCG (Besto India made with R_g= 200 Ω) and made connections as shown in figure 1. Then, we adjusted the Variable resistance R_1 and the value of low resistance r so as to produce ballistic throw within the scale. For R_e= 2 kΩ to 18 kΩ, the supply dc voltage (emf) was maintained at 16.2 V. This voltage was changed during measurements at R_e= 20 kΩ and at R_e= ∞. These changes in voltage was done so as to obtain noticeable deflection in the scale (see Table 1: the deflections at R_e= 20 kΩ and at R_e= ∞, marked * in Table 1, are not in proportion to the deflections at R_e= 2 kΩ to 18 kΩ). Such adustments in voltage does not affect the values of T and p as these quantities depend on external resistance R_e and other intrinsic constants of the galvanometer. However, low resistance r and other operational conditions remain the same throughout the experiment. We chose r = 0.01Ω. Deflections of the galvanometer were measured for different values of external resistance R_e following the procedure described in the previous sections and are presented in Table 1.

The readings are taken in cm from the linear scale. Obviously, these linear readings are proportional to the angular deflections of the coil.

The values of log decrements $\boldsymbol{\lambda}$ at each resistance were calculated using eq.(10) and then damping terms $\boldsymbol{p}$ were obtained using eq.(9). Thereafter, values of $\boldsymbol{c/I}$ were found from eq. (11). The values of $\boldsymbol{T}$, $\boldsymbol{\lambda}$, $\boldsymbol{p}$ and $\boldsymbol{c/I}$ for each resistance $\boldsymbol{R_e}$ are presented in Table 2. The error in the measurement of $\boldsymbol{\lambda}$ at $\boldsymbol{R_e}$= 2 kΩ was about 24% and hence it was ignored for further calculations. Errors in the individual measurements of $\boldsymbol{\lambda}$ for $\boldsymbol{R_e}$= 3 kΩ and above was well below 10 %.

On plotting a graph with $\boldsymbol{1/R}$ along x-axis and $\boldsymbol{p}$ along y-axis we obtained a straight line as shown in figure 2. From this graph we obtained slope, $\boldsymbol{G^2/2I}$ = 356 ± 12 Ω and intercept, $\boldsymbol{a/2I}$ = 0.006 ±0.002 (linear fit R^2 = 0.9955). The average value of $\boldsymbol{c/I}$, that is $\boldsymbol{(c/I)_{av}}$, was found to be 0.123 ±0.002. As expected, value of $\boldsymbol{c/I}$ in the open circuit condition is 0.123. Finally, critical damping resistance was obtained by using eq.(13). The experimentally measured value of critical damping resistance ($\boldsymbol{R_{cdr}}$) was found to be 833±49 Ω.

Table 1. Experimental data showing galvanometer deflections at different values of external resistances.

Resistance R_e (kΩ)	Time period T (s)	1st Throw θ_1 (cm)	θ_2 (cm)	θ_3 (cm)	θ_4 (cm)	θ_5 (cm)	θ_6 (cm)	θ_7 (cm)	θ_8 (cm)	θ_9 (cm)	θ_{10} (cm)	θ_{11} (cm)	θ_{12} (cm)
2	20.7	12.3	2.1	0.4	0.1								
3	19.2	9.9	3.1	1.2	0.2								
4	18.9	8.3	3.5	1.6	0.6	0.4							
6	18.3	6.4	3.6	2.2	1.1	0.7	0.3						
8	18.2	5.5	3.5	2.4	1.4	1	0.5						
10	18.1	4.8	3.4	2.4	1.6	1.2	0.8	0.5	0.4				
12	18	4.5	3.3	2.5	1.7	1.3	0.9	0.7	0.4				
14	18	4.3	3.3	2.6	1.9	1.5	1.1	0.9	0.6	0.5			
16	17.9	3.1	2.5	2	1.5	1.3	0.9	0.8	0.5	0.4			
18	17.9	3	2.4	2	1.5	1.3	1	0.8	0.6	0.5	0.4		
20	17.9	8.5*	6.7	5.6	4.5	3.7	2.9	2.5	1.9	1.6	1.3	1.1	0.8
Open	17.9	11.8*	11.4	10.7	10.3	9.7	9.4	8.7	8.5	8	7.7	7.2	7

Table 2. The values of Time period ($\boldsymbol{T}$), log decrement ($\boldsymbol{\lambda}$), damping term ($\boldsymbol{p}$) and $\boldsymbol{c/I}$ for each resistance R_e.

Resistance R_e(Ω)	Time period T(s)	Log decrement λ	Damping term p	c/I
2000	20.7	1.68	0.16	-
3000	19.2	1.11	0.12	0.122
4000	18.9	0.84	0.09	0.119
6000	18.3	0.58	0.06	0.122

8000	18.2	0.45	0.05	0.122
10000	18.1	0.37	0.04	0.122
12000	18.0	0.33	0.04	0.124
14000	18.0	0.27	0.03	0.123
16000	17.9	0.24	0.03	0.124
18000	17.9	0.22	0.03	0.124
20000	17.9	0.21	0.02	0.124
Open	17.9	0.05	0.01	0.122

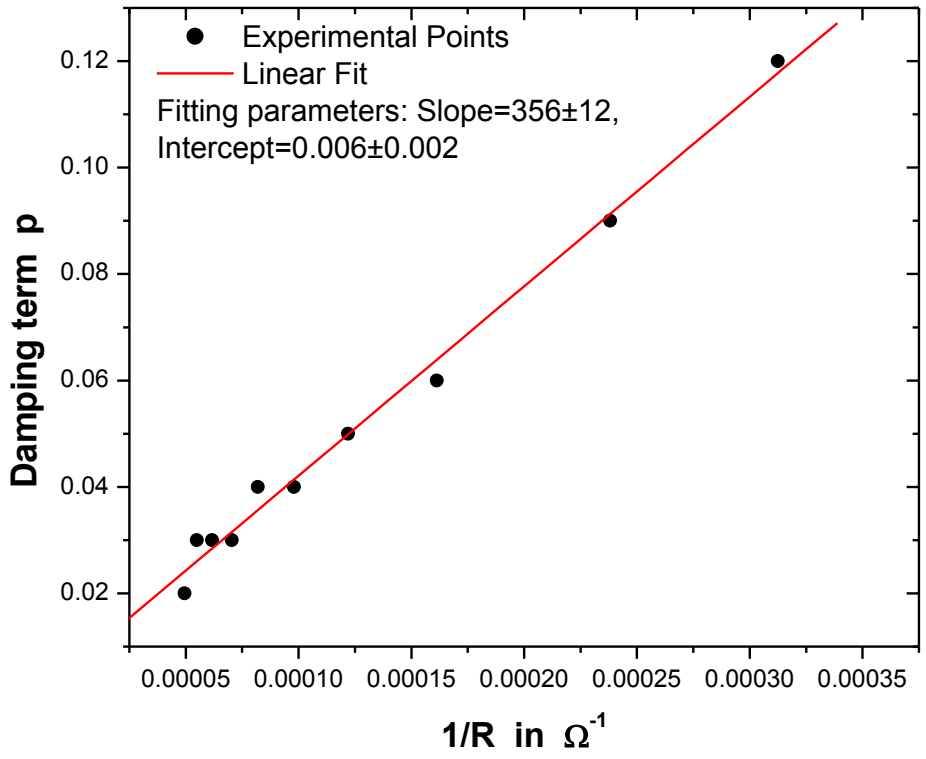


Figure 2. A plot with $\mathbf{1}/\boldsymbol{R}$ along x-axis and $\boldsymbol{p}$ along y-axis showing the variation of damping term p with external resistance $\boldsymbol{R_e}$. It is to be noted that the total resistance $\boldsymbol{R = R_g + R_e}$, where $\boldsymbol{R_g}$ is galvanometer resistance. The black solid points denote experimental data while the red coloured straight line is a linear fit.

5. Discussion

Even though most of the standard manufacturers of MCG provide the value of R_{cdr} , there are many manufacturers which provide no other data except the resistance of the Galvanometer. In such scenario it becomes important for the student to have, at least, a rough idea about the value of R_{cdr} so that the set up could be used as a BG. The method of finding critical damping resistance of an MCG presented in this paper is simple, straightforward and less time consuming. The working formulae used in this method have been analytically derived without making any approximations. In the illustrative experiment presented in this paper we have started our data from R_e = 2000 Ω onwards. At 2000 Ω, the coil makes only two oscillations before it comes to equilibrium position and the relative error here is about 0.25, hence this data was ignored for the calculations. We have only considered experimental data corresponding to R_e >2000 Ω for further calculations. We also observed that the value of *c/I* can be directly evaluated by substituting the values of *p* and *T* corresponding to open circuit in eq.(11). This way one can do away with the lengthy process of finding average value of *c/I* . The difference in values of *c*/*I* obtained from these two different technique is almost negligible. Although the number of oscillations made by the coil before it comes to rest at equilibrium position may be fewer at lower values of external resistance R_e it is advisable to include these readings as the variation of time period and logarithmic decrements are prominently seen at lower resistances. Besides knowing the value of CDR this method also gives an idea about the relative values of G , a and c . The overall accuracy in the determination of the value of R_{cdr} in this experiment is within 5 %. The main focus of the present work is the presentation of the working formula and methodology. The experimental part has been included in the present work primarily to give a fillip to the understanding of the underlying theory and procedure.

References:

[1]O.M. Stewart, The Damped Ballistic Galvanometer, Phys. Rev. 16,158(1903)

[2]W. P. White, Sensitive Moving Coil Galvanometers, Phys. Rev. 19, 305(1904)

[3]W. Jaeger, Über das Drehspulengalvanometer, Annalen der Physik 326, 64(1906)

[4]W. P. White, Every-day Problems of the Moving Coil Galvanometer, Phys. Rev. 23, 382(1906)

[5]F. Wenner, General Design of Critically Damped Galvanometers, Bulletin of the Bureau of Standards 13, 211(1916)

[6]R. N. Rai, Moving Coil Galvanometer and Critical Damping, American Journal of Physics 12, 151 (1944);

[7] G. W. Rodeback, The Moving Coil Ballistic Galvanometer Revisited, Rev. Sci. Instrum. 43, 837 (1972)

[8]Feweks J.H. and Yarwood J. *Electricity and Magnetism Volume I,* University Tutorial Press Ltd., London,1965

Novel method for determination of contact angle of highly volatile liquids

V Madhurima and K Nilavarasi

Department of Physics,
Central University of Tamil Nadu,
India-620015.

nilavarasikv@gmail.com

Submitted on 07-09-2021

Abstract

In this paper, a novel method for the measurement of equilibrium contact angle of highly volatile binary liquids is proposed. The proposed computational method, which combines finite element method and energy equilibration, calculates the solid-liquid contact area, which can then be used to estimate the equilibrium contact angle. Using the proposed approach, the contact angles of binary-liquid droplet on a microgrooved and smooth polycarbonate substrate were calculated. The proposed method is found to be an efficient tool for finding the contact angle of all liquids (both volatile and non-volatile).

1 Introduction

Wettability becomes crucial for many industrial and scientific applications such as painting/coating [1, 2, 3, 4], surface chemistry [5], oil recovery [1, 2, 3, 4] and so on. Wettability describes the balance of three inter-facial interactions namely, solid-liquid (γ_{sl}), liquid-vapor (γ_{lv})and solid-vapor (γ_{sv}) inter-facial interactions [6, 7, 8]. The balance between such three interactions are expressed by Young's equation [9, 10, 11],

$$\gamma_{lv} cos\theta = \gamma_{sv} - \gamma_{sl} \tag{1}$$

where θ is the contact angle. Although the surface tension of the liquid can be measured experimentally with satisfactory accuracy, the solid-liquid inter-facial tension cannot be measured directly, and therefore the wettability is usually described by the contact angle, which is the angle formed between the solid-liquid interface and the liquid-vapor interface. The contact angle is measured experimentally using the standard sessile drop method, in which the camera focuses on a liquid droplet placed over a solid substrate and the geometry of the droplet is used to obtain the contact angle [12].

Apart from sessile droplet method, there are also other methods of measuring contact

angle directly which includes ”tilting plate” method [13, 14], captive bubble method, etc., [15, 16, 17]. Although the measurement of contact angle from these methods is relatively straightforward, there are issues that require attention while using volatile liquids: 1. The inherent inaccuracy of the direct measurement techniques and 2. simultaneous variation in the contact area and contact angle of the liquid over the solid surfaces [18]. There have been numerous studies reporting contact angles for a variety of liquids, binary systems, etc., In most such studies, the liquids used are less volatile and the binary system used contains water as one of its moiety [19].

In the present paper, an attempt is made to address the problem of measuring contact angle of highly volatile liquids. Here, equilibrium droplet shape of highly volatile binary liquids on the horizontal smooth and constrained surfaces are simulated to obtain the contact angle of binary liquids on solid surfaces. This 3D-drop shape model is used to numerically analyze the shape, contact area and contact angle of the liquid droplets over the solid surface. The effect of variation of surface tension and surface roughness on the drop shape and apparent contact angle is examined. The liquids used in the present study include methanol, ethanol and different concentrations of the ethanol-methanol binary system. The simulated results are validated with the experimentally obtained data. The present study shows that the proposed method can be an efficient tool for finding the contact angle of all liquids (both volatile and non-volatile).

The contact angle measurements are also useful in determining the size of the pores formed as a result of self-assembly of condensed liquid droplets. Since, pore size is a measure of diameter of a triple phase contact line. In the present study, an attempt is also made to compare the experimental pore size of the self-assembled droplet patterns with the simulated droplet size.

2 Simulation Details

The equilibrium shape of a liquid drop is achieved through energy minimization. In the present study, the finite element method and gradient descent method is used to evolve the surface toward minimal energy using *Surace evolver* [20, 21]. In this the liquid droplet equilibration is achieved by minimizing various energies namely, surface tension, gravitational energy, etc., involved in the defined system. The initial geometrical parameters, energies and constraints involved in the system are given as inputs and the program minimizes the total energy of the system by modifying the surface geometry according to the defined parameters and constraints. The droplet size of the binary liquid obtained from simulated stable equilibrium shape of various liquid droplets over specific smooth and constrained surfaces are compared with those from experiments.

In the present work, the gravitational effect

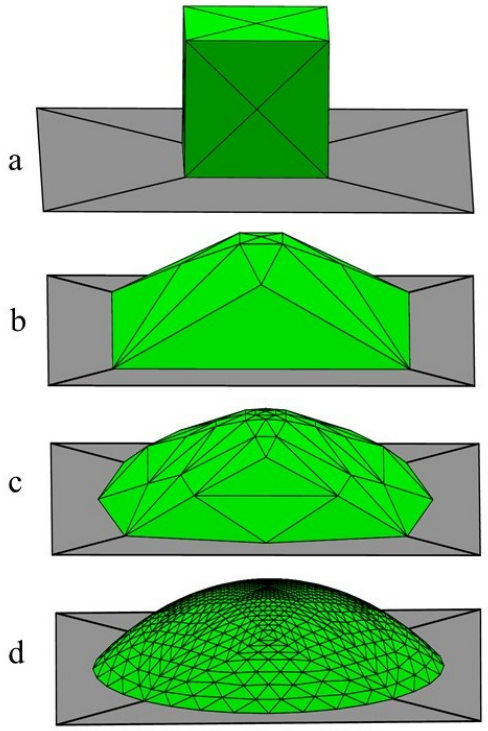


Figure 1: System evolving from initial geometry. a) Initial drop b) After 100 refinements c) After 200 refinements and d) Final shape of the droplet.

is negligible and hence do not influence the results significantly. Therefore, gravitational energy was not taken into account. The initial geometry and the shape after successive evolutions are shown in Figure 1. The free energy of the system is expressed as [22],

$$G/\gamma_{lv} = A_{lv} - \int\int \cos\theta dA \qquad (2)$$

and the contact angle (θ) is defined by Young's equation [9, 10, 11]. Here in equation 8, γ_{lv} refers to surface tension of the liquid and A_{lv} refers to the liquid-vapor contact area. The drop volume specified in this work is 9 μl. The bottom face of the droplet is constrained to move and this boundary condition is considered to be responsible for obtaining the shape of the droplets. For the sake of convenience, the constraints are specified to the edges which defines the three phase contact line. The successive refinements and steps concerning energy minimization computes the equilibrium shape of the system. Figure 1 gives the illustration of the steps involved in the *Surface Evolver* simulation.

3 Results and Discussion

3.1 Smooth surfaces

Measuring the contact angle of the highly volatile liquid is experimentally difficult [5, 15]. Hence, in the present study, *Surface Evolver* is used to model the wetting behavior of the ethanol-methanol binary liquid drop on the smooth surface. For this solid-liquid inter-facial tension calculated from experiments is used an input parameter instead of conventionally used contact angle. The variation of solid-liquid inter-facial tension for various concentration of methanol is shown in Figure 2.

The shape of the surfaces after the minimization of surface free energy are shown in Figure 3 and 4. It is observed from the figure that the variation of concentration of binary liquid affects the width and height of the liquid drops on the smooth surface. Hence the shapes calculated from *Surface Evolver* helps in understanding the variation of pores shapes with varying concentration of methanol [23].

The solid-liquid inter-facial tension for all concentration of ethanol-methanol binary sys-

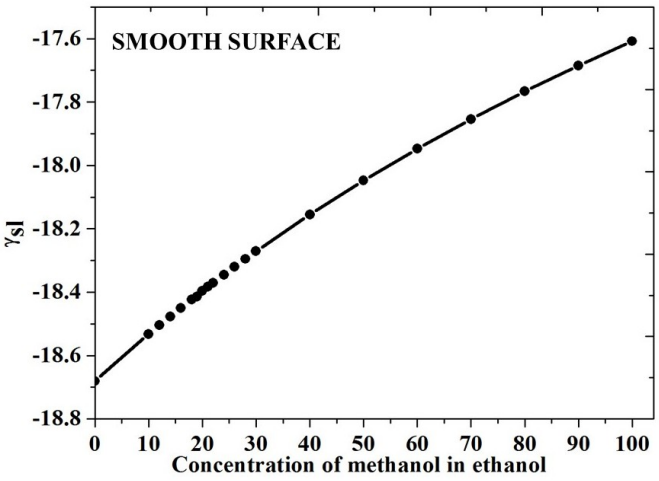


Figure 2: Variation of solid-liquid inter-facial tension of smooth surface for various concentration of methanol in ethanol.

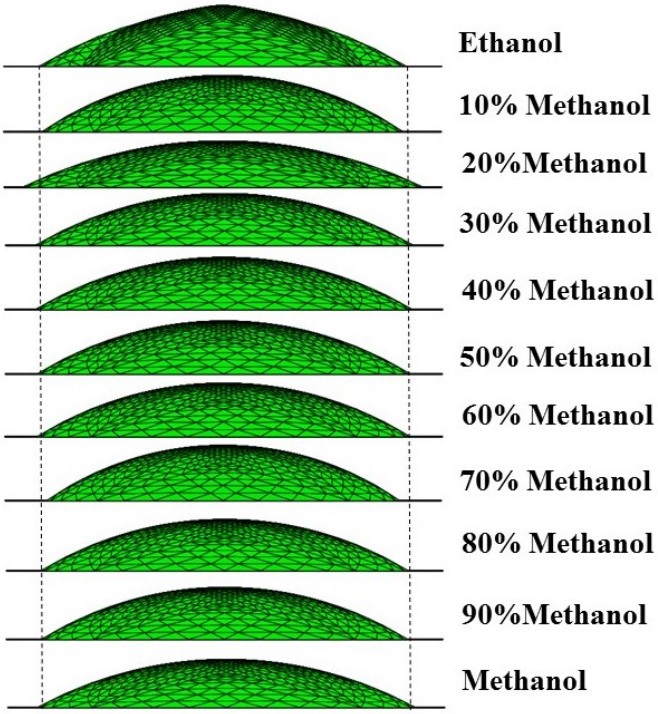


Figure 3: Variation of droplet diameter on smooth surfaces simulated using *Surface Evolver*.

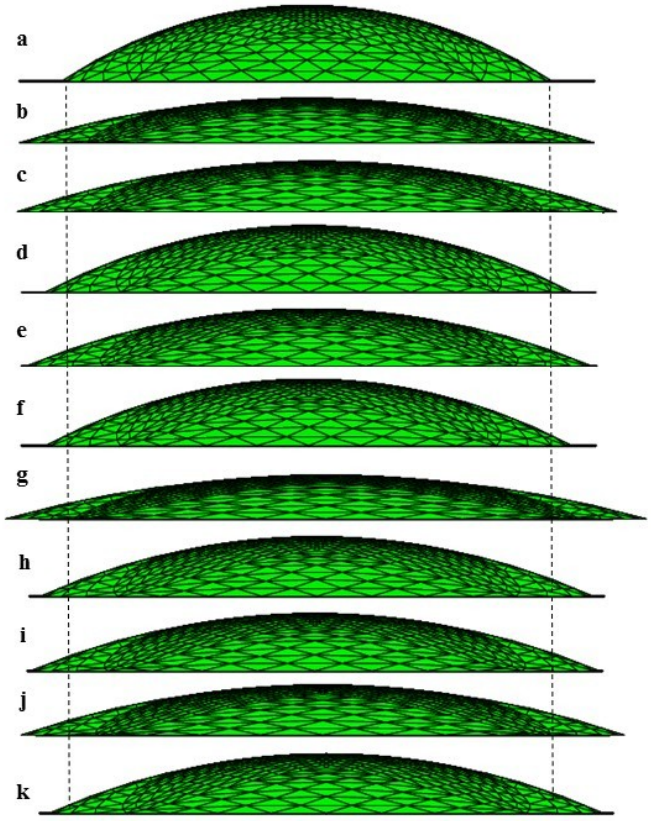


Figure 4: Variation of droplet diameter on smooth surfaces (From the top a)12% b) 14% c) 16% d)18% e) 19% f) 20% g) 21% h)22% i) 24% j) 26% k) 28% of methanol in ethanol-methanol binary system).

tem is calculated following ideal Raoult's law [24, 25] and using the thus obtained values, the drop shapes are simulated. The results of *Surface Evolver* simulation are compared with the experimentally observed profile of pore shapes. The simulation results show a similar trend as observed in experimental results [23]. The complex variation at low concentration of methanol as seen in experiments is also observed in simulation results. To further understand the correlation between the concentration of methanol and the pore size, diameter of the droplets are cal-

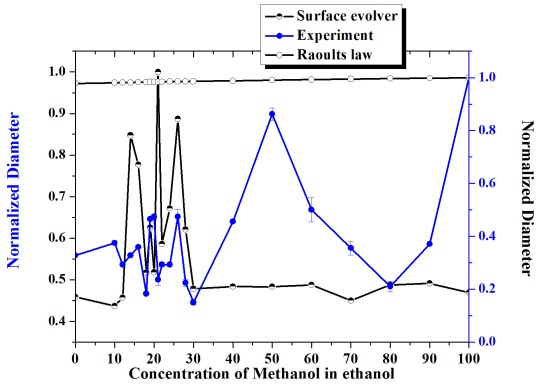


Figure 5: Comparison of normalized diameter on smooth surface calculated from simulation and experiments.

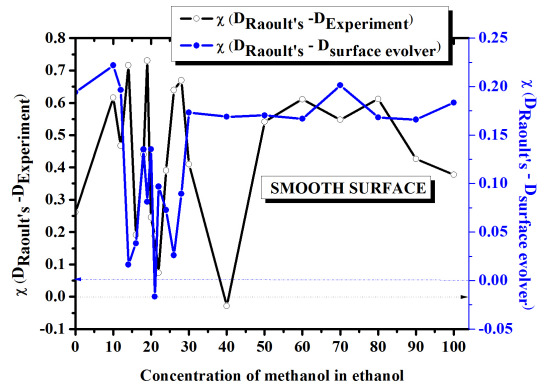


Figure 6: Deviation in normalized experimental pore diameter and simulated droplet diameter from the ideal case (For smooth surface).

culated. From the obtained diameter of drops, normalized diameter is calculated. The normalized diameter versus concentration of methanol is also plotted (shown in Figure 5). Comparison is made between the normalized diameter calculated from simulation and the normalized pore size obtained from experiments and is shown in Figure 5.

Further, the deviation in diameter obtained from experiment and simulation from the ideal case (diameter obtained from simulations using inter-facial tension,which is calculated from Raoult's law) is evaluated and is shown in Figure 6. It is observed that except at the concentration of 21% and 40% of methanol, all other concentrations showed a positive value of the deviation in diameter. This concentration 21% of methanol is the concentration where the complexity in intermolecular interactions is observed [26] and 40% of methanol is the concentration, where the presence of strong hydrogen bonds have been reported [27, 26].

The difference between the simulation and experimental diameter is also calculated and is shown in Figure 7. Positive deviation is observed for all concentration of methanol except at 21% of methanol. This shows that the simulated results are in good agreement with the experimental data i.e., the relative error between experimental and simulated values are small. The negative value at 21% occurs in the region where the complex behavior in inter-molecular interactions is observed. For the sake of comparison, the liquid-vapor inter-facial tension obtained from experiments and calculated from Raoult's law is plotted and shown in Figure 8. It is clear from the figure that the ethanol-methanol binary system is far from ideal.

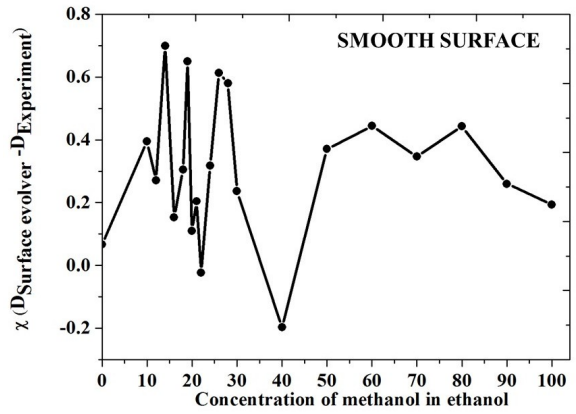


Figure 7: Difference between normalized diameter (on smooth surface) obtained from simulation and experiment.

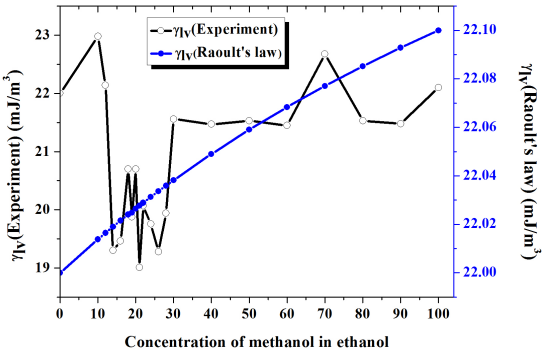


Figure 8: Surface tension of liquids obtained from experiments and Raoult's law.

The total energy and total area of the system for various concentration of methanol are calculated from the simulations and are shown in Figure 9. The total energy and total area show a similar trend as seen from experimental result at the lower concentration of methanol in ethanol-methanol binary system.

Hence in conclusion, the simulation results are found to corroborate the experimental findings, thus proving to be a tool for calculating the contact angle of highly volatile liquids for which experimental measurement of contact angle is difficult.

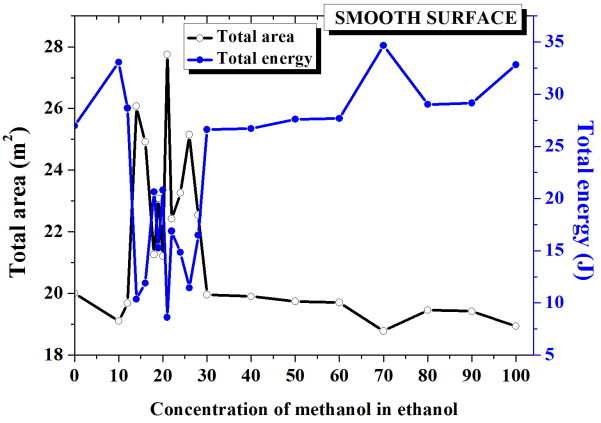


Figure 9: Variation of total energy and total area of the system (liquid droplet on smooth surface) for various concentration of methanol in ethanol.

3.2 Constrained surfaces

The parallel striped surfaces of equal distance between the stripes is simulated and are used as constrained surfaces. These parameters of these surfaces are specified such that they match properties of the parallel grooved surfaces as used in the self-assembled droplet pattern formation (whose pore sizes are compared here in the present work) [23]. The solid-liquid inter-facial tension for constrained surfaces are calculated from experiments and used as an input parameter instead of conventionally used contact angle. The variation of solid-liquid inter-facial tension for various concentration of methanol is shown in Figure 10.

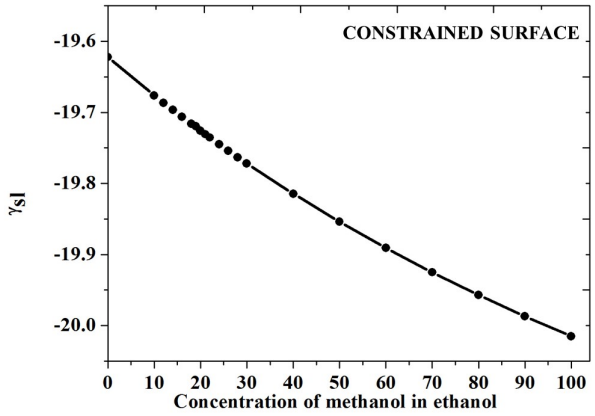


Figure 10: Variation of solid-liquid inter-facial tension of constrained surfaces with various concentration of methanol in ethanol.

The shape of the droplet on constrained surfaces after the energy minimization are shown in Figure 11 and 12. It is observed from the figure that the variation of concentration of binary liquid affects the width and height of the liquid drops on the constrained surface. Hence the shapes calculated from simulations helps in understanding the variation of pores shapes with varying concentration of methanol. It is also observed that the constraints on the surface has significant influence in the wetting behavior of the surface.

Similar to smooth surfaces, the solid-liquid inter-facial tension for all concentration of ethanol-methanol binary system is calculated following ideal Raoult's law and using the obtained values the drop shapes are simulated. The results of the simulation are compared with the experimentally observed profile of pore shapes on constrained surfaces. The simulation

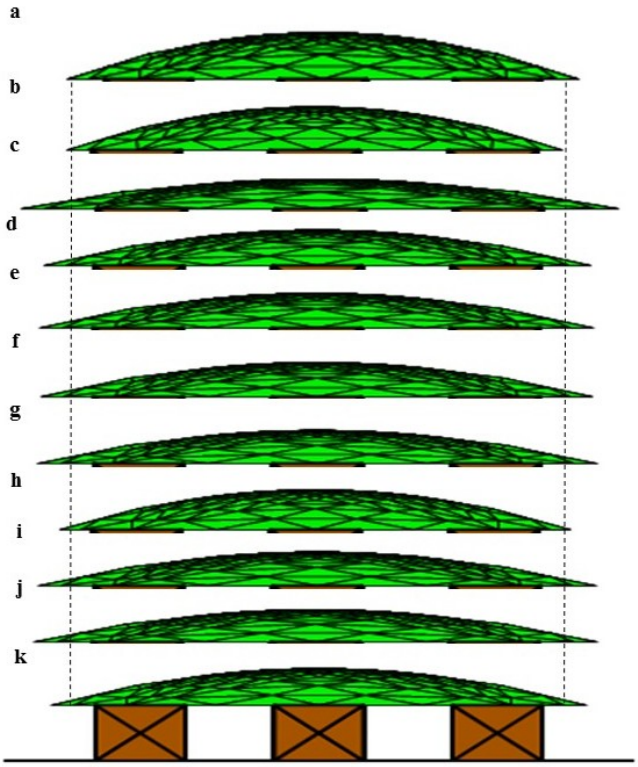


Figure 11: Variation of droplet diameter on constrained surfaces (From the top a) 0% b) 10% c) 20% d) 30% e) 40% f) 50% g) 60% h) 70% i) 80% j) 90% k) 100% of methanol in ethanol-methanol binary system).

results show a similar trend as observed in experimental results. The complex variation at low concentration of methanol is also observed in the simulation results. To further understand the correlation between the concentration of methanol and the pore size, diameter of the droplets are calculated. From the obtained diameter of drops, normalized diameter is calculated. The normalized diameter versus concentration of methanol is also plotted (shown in Figure 13). Comparison is made between the normalized diameter calculated from simulation and the normalized pore size obtained from experiments

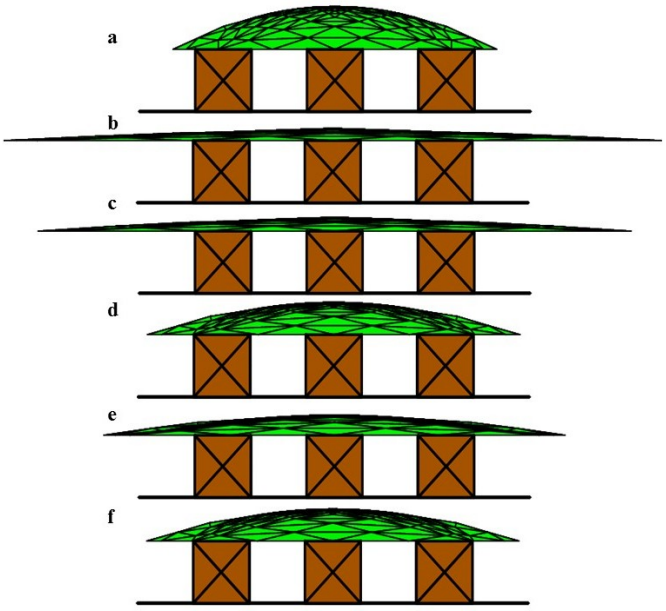


Figure 12: Variation of droplet diameter on constrained surfaces (From the top a) 12% b) 14% c) 16% d) 19% e) 20% f) 21% of methanol in ethanol-methanol binary system).

and is shown in Figure 13. Further, the deviation in diameter obtained from experiment and simulation from the ideal case (diameter calculated from simulations using inter-facial tension calculated from Raoult's law) is calculated and is shown in Figure 14. Positive deviation is observed for all concentration of methanol except at 21% of methanol and 100%, where the complexity in properties is observed. At certain concentrations, the datas obtained from experiments and simulations doesnt correlate. This is due to the presence of chloroform used in the experiments which is not actually taken into account in simulations.

The difference between the simulation and experimental diameter for constrained surfaces

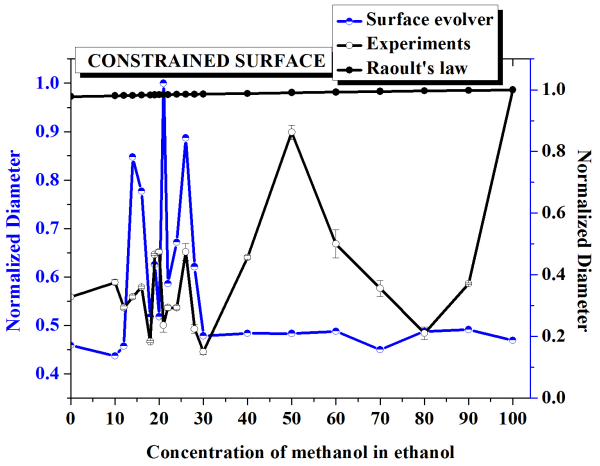


Figure 13: Normalized diameter of droplet on constrained surfaces obtained from simulation and experiment.

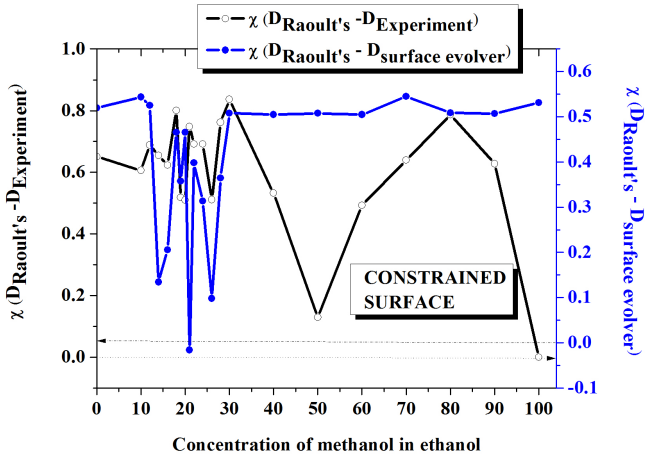


Figure 14: Deviation in experimental pore diameter and simulated droplet diameter from ideal case (constrained surfaces).

is also calculated and is shown in Figure 15. It is observed that the deviations is observed to be positive for low concentration of methanol. The higher concentration shows a negative deviation, which may be due to the possibility of complex interactions as observed in [23, 26]. For the sake of comparison, the liquid-vapor

inter-facial tension obtained from experiments and calculated from Raoult's law is plotted and shown in Figure 8. It is clear from the figure that the ethanol-methanol binary system is far from ideal.

The total energy and total area of the constrained system for various concentration of methanol are calculated from the simulations and are shown in Figure 16. The total energy and total area show a similar trend as seen from experimental result at the lower concentration of methanol in ethanol-methanol binary system.

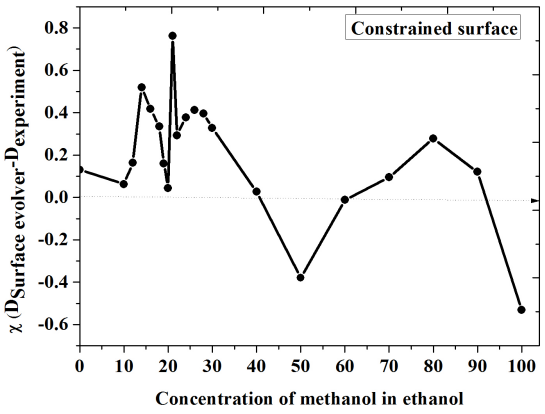


Figure 15: Difference between the normalized simulated and experimental diameter for constrained surfaces.

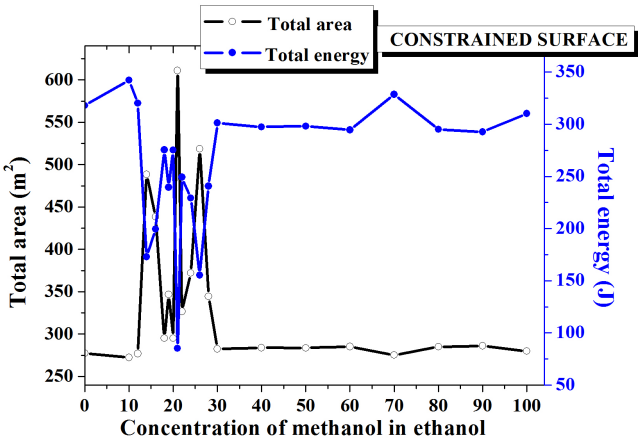


Figure 16: Variation of total area and total energy of the system (liquid droplet on constrained surfaces) .

Similar to the simulations obtained for smooth surfaces, the simulated results of constrained surfaces is found to corroborate the experimental findings. The proposed method is proved to be a tool for calculating the contact angle of highly volatile liquids for which experimental measurement of contact angle is difficult.

4 Conclusion

The contact angle of highly volatile liquid droplets on horizontal smooth and constrained surfaces are calculated. This is performed by simulating the equilibrium liquid droplet shape on the specified surfaces and numerically analyzing the contact area and contact angle of the liquid droplets on the surfaces. The effect of variation of surface tension and surface roughness on the drop shape and apparent contact angle are examined through the simulating the droplets of methanol, ethanol and various concentration of ethanol-methanol binary system. The simulated results are validated with the experimentally obtained data. The present study shows that the proposed simulation method can be an efficient tool for finding the contact angle of all liquids (both volatile and non-volatile).

References

[1] K.N. Prabhu, P. Fernandes, G. Kumar, Mater. Des. 2, 297 (2009)

[2] X. Zhao, M.J. Blunta, J.J. Yao, Pet. Sci. Technol. Eng. 71, 169 (2010)

[3] Y.Q.Wang, H.F. Yang, Q.G. Hang, L. Fang, S.R. Ge, Adv.Mater. Res. 154-155, 1019 (2010)

[4] M. Sakai, T. Yanagisawa, A. Nakajima, Y. Kameshima, K. Okada, Langmuir 25, 13 (2009)

[5] H. B. Eral, D. J. C. M. 't Mannetje, and J. M. Oh., Colloid Polym. Sci., 291, 2:247-260, (2013).

[6] Robert J. Good, Journal of Adhesion Science and Technology, 6:12, 1269-1302, (1992).

[7] Gregory J. Merchant, Physics of Fluids A: Fluid Dynamics, 4, 477 (1992).

[8] L. M. Hawking, Physics of Fluids, 7, 2950 (1995).

[9] T. Young, Philos. Trans. R. Soc. Lond. 95, 65 (1805)

[10] A. M. Dupre, Theorie Mechanique de la Chaleur, Gauthier-Villars, Paris, 369, (1869).

[11] M. E. Schrader, Langmuir, 11, 3585-3589, (1995).

[12] W.C. Bigelow, D.L. Pickett, W.A.J. Zisman, Coll. Sci. 1, 513 (1946)

[13] G. MacDougall, C. Ockrent, Proc. R. Soc. 180A, 151 (1942)

[14] C.W. Extrand, Y. Kumagai, J. Colloid Interface Sci. 184, 191 (1996)

[15] K. L. Mittal and R. Jaiswal, Particle Adhesion and Removal, Scrievener publishers, (2014).

[16] A.F. Taggart, T.C. Taylor, C.R. Ince, Trans. Am. Inst. Min. Metall. Pet. Eng. 87, 285 (1930).

[17] W. Zhang, M. Wahlgren, B. Sivik, Desalination, 72, 263 (1989)

[18] Fowkes, Contact Angle, Wettability, and Adhesion Advances in Chemistry; American Chemical Society: Washington, DC, (1964).

[19] Z.Zhang, W. Wang, A. N. Korpacz, C. R. Dufour, Z. J. Weiland, C. R. Lambert, M. T. Timko, Langmuir, 35: 38, 12317 (2019).

[20] K. A. Brakke, The Surface Evolver, Exp. Math. 1, 141, (1992)

[21] K. A. Brakke, 2008, Surface Evolver manual: version 2.30, http://www.susqu.edu/brakke/aux/downloads/manual2

[22] Y Chen, B He, J Lee, N A Patankar, Journal of Colloid and Interface Science, 281, 458-464, (2005)

[23] K. Nilavarasi and V. Madhurima, Eur. Phys. J. E, 41: 82, (2018).

[24] D. A. McQuarrie, J. D. Simon, University Science Books, (1997).

[25] E. B. Smith, Clarendon Press, Oxford, (1993).

[26] K. Nilavarasi, Thejus R Kartha and V. Madhurima, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 188, 301-310, 2018.

[27] V. H. Alvarez, S. Mattedi, M. Iglesias, R. Gonalez-Olmos and J. M. Resa, Physics and Chemistry of Liquids, 49:1, 52-71, (2011).

Advances and challenges in VLSI designs for nanotechnology

Amit Kumar

Department of Electronics
Bhaskaracharya College of Applied Sciences
University of Delhi, South West Delhi 110075, India.

amit.kumar@bcas.du.ac.in

Submitted on 15-07-2021

Abstract

Since the advent of nanotechnologies, Moore's law has been challenged through advances in molecular science and fabrication techniques, as the feature size reached the molecular scales where classical physics is impossible to explain and quantum mechanics takes over. The present trend is toward VLSI design with CMOS as it is the most efficient at lower prices and offers higher density. The scaling theory is used when VLSI designs move from old fabrication technology to new fabrication technology resulting from shrinking circuit sizes. CMOS technology now no longer follows scaling theories since we have reached the sub-nanometer regime. A short channel effect occurs when channel lengths of aggressively scaled MOSFETs are comparable to the sum of the source and drain depletion layer thicknesses. The problem of short-channel effects can now be solved by proposing alternative structures and materials. In this paper, we discuss a few of these alternatives along with the current situation (covering the past decade) and possible future trends. This paper summarizes the VLSI design developments in last decade and near future opportunities due to advent in nanotechnology.

Key-words: *Microelectronics, VLSI designs, Nanotechnology, Review article*

Background

Moore's empirical observation [1] of integrated circuit density doubling every year is being followed since the 1970s and 1980s. It was with the advent of microprocessor-equipped products like the Hewlett Packard calculator, the Apple II computer, and the IBM PC that integrated circuits (ICs) became a mainstream technology in early 60s. There has been a steep increase in the demand for these products because of the fierce competition among manufacturers since then. Ever since, the aim has been to fit more capable chips into small packages. The last five decades have seen transistor counts per microprocessor doubling every two years, resulting in an obvious increase in the processing power at high speeds, a device can offer. The techniques of integrated electronics are almost imperative in any system of operation, from basic field agriculture to advanced space programs. As the performance of microprocessors improves, so do the components of their circuit, which must be scaled down so that a greater number of elements can be packed in a smaller area. Thus, electrons can move more quickly, which leads to higher processing speeds. Since 2004, the clock speed and the rate at which instructions are executed have been slowed down by heat.

Since the late 1950s, integrated electronics has encompassed various technologies known as microelectronics. The thrust has been on increasing

the complexity in short spaces and with the minimum amount of weight while miniaturizing electronic components. The IC revolution in the last three decades is primarily due to the steady downward trend of lithographic design rules. The downscaling process involves the refinement of photolithography, which is the key technique for etching microscopic elements on silicon surfaces [2]. Earlier efforts had been made to achieve higher resolution miniaturization by reducing all other circuit components to the same size as the transistor. ICs, therefore, had to be scalable, and the new transistor structure that could achieve this was a MOSFET (Metal Oxide Silicon Field Effect Transistor) [3]. One of the most important issues of microprocessor design is the power consumption, which should always be carefully considered with a goal of making the device portable while delivering competitive performance [4]. The low-power processor and system-on-a-chip are two current trends that are supported by complementary-MOSFET (CMOSFET) technology. As manufacturing techniques have advanced, the theoretical limitations of design have arisen as designing such a low scale device has become increasingly challenging. The numerous short-channel effects, such as threshold voltage roll-off, subthreshold current, drain-induced barrier lowering, off-state current, hot carrier effect, gate-induced drain leakage, mobility degradation, *etc*., put an upper limit on VLSI chip designs. After 2005, there have been considerable efforts made by designers to break these threshold limits, but now it seems we need to find more viable solutions. Designing an aggressive scale electronic chip is responsible not only for the advent of technologies, but in turn, for the advancement of human civilization.

The present time may be seen as a transition time between two technologies, traditional VLSI design-based alternatives and emerging alternatives as suggested by the advent of nanotechnology. This is the time to review current situation as to why new technologies are needed and how current ones are limited. The purpose of this review article is to systematically examine what solutions exist to break Moore's law threshold limits of current VLSI designs due to advances in nanotechnology, so that it can also continue in the future.

2. Current trend in VLSI designs

The development of manufacturing technologies has registered the fundamental limits of silicon CMOS processing and the timeline of the IC's feature size can be seen in figure 1. A 32nm Toshiba NAND flash memory chip with 32GiB memory capacity was commercially released in 2009. At Micron Technology, Gurtej Singh Sandhu developed pitch doubling patterns during the mid-2000's that were used in 32nm technology during that time period [5]. Due to Sandhu's technology, extreme device downscaling became possible, and he is recognized as one of the seven most prolific inventors of all time. Sandhu, who graduated from the Indian Institute of Technology in Delhi, developed an innovative method of coating microchips with titanium without exposing them to oxygen. When he reported this, he didn't realize that this process would be the basis of most chip production in the future. In the early 2010s, Intel began selling its 32nm processors, called Clarkdale processors (which had previously adopted 45nm technology), based on Westmere architecture (which was derived from Nehalem). Traditionally, Intel concentrated on using die shrinks to boost performance at regular intervals through lithographic tick-tock modeling, *i.e.* the microarchitecture (tock) follows the die shrink (tick) for improved performance with the same microarchitecture [6]. Die shrink modeling was so beneficial that it decreased the current used by each transistor while maintaining the same clock frequency of the chip, therefore resulting in less power consumption, and increasing clock rate headroom at a lower cost.

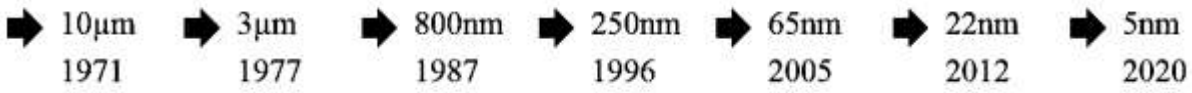


FIG. 1: Timeline for IC's feature size (past 50 years).

However, there were many difficulties in implementing the tick, and the decreasing procedures for node size seemed to suffer from quantum complexity in the end. The tick-tock process has been replaced by Process-Architecture-Optimization (PAO). Ambarella Inc., a pioneer in low-power, high-definition video compression, and image processing semiconductors, demoed the A79 system-on-chip (SoC) using 32nm technology, on September 26, 2011. Based on 16-bit DDR3 DRAM, this system has a low power consumption and low cost. It has also enabled the creation of a new generation of HD video using digital still cameras. Later, AMD released a 32nm processor, Bulldozer based on Silicon on Insulator (SoI) technology with two CPU cores per module up to four modules ranging from quad-core design.

Intel's Ivy Bridge CPU family (2nd generation of Intel Core processors, in i7 & i5 series) is based on the 22nm node process, which followed 32nm MOSFET (CMOS) node technology [7]. The success of Intel's 22nm process that went into Ivy Bridge gives them a clear advantage over other companies. The processors especially save power because of transistor leakage; the thermal design power (TDP) for Ivy Bridge desktops was 77W, whereas Sandy Bridge's 32nm predecessor consumed 95W. In 2006, the International Technology Roadmap for Semiconductors (ITRS) updated the assumption that the oxide thickness would not decrease below 0.5nm, i.e. about twice the size of silicon atom, which was the expected value at 22nm lithographic designs. There was a concern that this could disrupt Moore's Law being followed through CMOS scaling [8]. A team of international researchers led by Digh Hisamoto from Japan's Hitachi Central Research Laboratory demonstrated 22nm technology at UC Berkeley in 1998. Researchers in this group reported a new self-aligned double-gate MOSFET structure called the FinFET by using boron-doped Ge/sub 0.6/Si as gate materials. The quasilinear structure that was fabricated using conventional planar MOSFET processes was a new variant of the vertical double-gate MOSFETs. Several companies, including Freescale, IBM, ST Microelectronics, Toshiba, and College of Nanoscale Science and Engineering, announced in 2008 that they had developed and manufactured 22nm SRAM cells, built over traditional six transistor designs on 300mm wafers. This structure suppressed short channel effects, which led to new technological advances when Intel introduced the i5 and i7 series of microprocessors in 2012. The reverse engineering of these devices revealed that Intel's 22nm Ivy processor showed an almost triangular-shaped cross-section that was referred to as tri-gate transistors by the manufacturer [10]. In the simulation studies with light doping profiles in the channel, it was found that there is no significant difference between rectangular and trapezoidal (nearly triangular) FinFETs. The triangular section was very different from the rectangular idealized sections Intel had previously shown in 2011. It is not clear whether this design was intended to have a critical impact on electron mobility or yield. The rectangular fin design showed better results in terms of short channel effects, although it is unclear whether it was planned or due to the finite-element technology involved in terms of fin etching. A multigate FET (MuFET) adds more than one gate to a single FinFET structure, allowing for further improvement. These devices used independent gate electrodes, and hence they are also known as multiple independent gate FETs (MIGFET) and gate all around FETs (GAAFET). All these have non-planer geometry and have been proven to successfully extend Moore's law with 3D transistors. The development of these new device structures is credited to Toshiba, Samsung, Intel,

UC Berkeley, IBM, Hitachi, AMD, *etc.* FinFETs were the most complex devices in recent times when they were first introduced in 22nm technology. The successor of the 22nm node is the 14nm that Intel announced in August 2014. The 14nm microarchitecture is used for the Core M processors used in systems, as reported previously [11]. First 14nm technology is used in volume production using second-generation tri-gate (FinFET) transistors. It claimed the industry's leading performance, power density, and cost power transistor. The technology promises to be used for high-performance applications such as advanced servers, personnel computing devices, and the Internet of Things. On September 5, 2014, Intel released its first Broadwell-based processors, the M-5Y10, M-5Y10a, and M-5Y70. In such applications as cloud/datacenter servers, high-end mobile processors, advanced driver assistance systems, IoT edge computing, *etc.*, 14nm FinFET is proving to be an ideal technology.

The first 10nm gate oxide was demonstrated in 1987 when Davari *et al.* reported a submicron tungsten gate for MOSFETs [12]. An international team of researchers at UC Berkeley published the first FinFET with a gate length of 10nm with a fin length of 12nm, in 2002 [13]. Structures described here outperform short-channel effects in silicon MOSFETs of any single layer. They demonstrated the operation of CMOS FinFET inverters based on the structures. Samsung was the first company to mass-produce a multilayer cell flash memory chip using a 10nm process. Later in October 2016, Samsung Electronics started mass producing 10nm SoC chips, and in April 2017 it began producing the Galaxy S8 smartphone, which uses a 10nm processor. This short-lived node was mainly designed for Apple processors during 2017-18, but was later replaced by 7nm during 2018. The notion of a node in regards to gate length, metal pitch, or gate pitch is becoming blurred. For example, Intel's 10nm process is similar to Global Foundries' 7nm process, and Samsung's 10nm process is somewhere between Intel's 14nm and 10nm processes in terms of transistor density. As in terms of transistor density, transistor size may not be as important as transistor density (number of transistors per square millimeter) in that a smaller transistor may not necessarily improve performance. Taiwan Semiconductor Manufacturing Company (TSMC) announced production of 256Mbit SRAM memory chips using a 7nm process in April 2017. A 7nm manufacturing process is also planned for Intel, but the company hasn't outlined what its expected value may be. Intel invested $7 billion for the next-generation semiconductor factory in Arizona to meet Moore's law. This new generation of microprocessors will enable Intel to run the latest computers, data centers, sensors, and other high-tech devices. It will enable technologies such as artificial intelligence, more sophisticated vehicles, transportation services, and medical treatment, among others. The current design trend is based on MuGFET and GAAFET technologies, which are not yet commercially available. Intel announced in December 2019 that it would begin producing 3nm chips in 2025 and 1.4nm chips in 2029. It will be a challenge to transition to 3nm designs in the coming two years, around 2023, with some roadblocks to overcome.

3. Current challenges in VLSI designs

From the early 1970s until the present, MOSFETs were successfully used in IC fabrication. From the early 1980s to the end of the last decade, CMOS-VLSI designs were among the favorite designs of design engineers due to their easier scaling down benefit [16]. With smaller CMOS-VLSI feature sizes, leakage power and propagation delays became greater challenges. In order to achieve better performance tradeoffs, the threshold voltages were continuously scaled down, which resulted in an exponential increase in static power dissipation due to the increase in subthreshold leakage current. Initially, the reduction of threshold voltage was deemed beneficial to speed, however, it affected noise immunity. At deep submicron levels, static power dissipation becomes almost proportional to switching dynamic power dissipation. Since the world has become increasingly reliant on battery-operated portable electronics, manufacturers have

focused on power-saving electronic designs that do not need bulky cooling enclosures. Moore's law has been the guiding principle for cost-effective miniaturization of electronics for decades. Based on the Dennard scaling principle, Moore's law has been successfully followed since the channel length scaled-down, and the oxide thickness scaled proportionately, to achieve strong capacitive coupling between the inversion channel and gate terminals, relative to the other transistor terminals. In order to meet the demands of miniaturization as per Moore's estimations, the International Technology Roadmap for Semiconductors (ITRS) was launched in 2000 [17]. In late 2000, it was confirmed that conventional scaling rules could no longer be applied to 32nm CMOS technology and beyond due to unsurmountable challenges. FinFET is a novel device architecture designed to provide greater gate control over traditional FET devices. To get additional control over the channel current, it uses a multi-gate transistor architecture. Increased gates enabled better electrostatic control, but at the cost of increased complexity. In the last decade, FinFETs were reported on both bulk silicon and silicon on insulator substrates. The un-doped channel has mitigated variability risks and reduces leakage current by a factor of four when compared to 32nm planner CMOS technology [18]. Thus, MuFET enabled VLSI technology at nanometer nodes and FinFET enabled continuous scaling down due to its ultrathin body structure. It was FinFET's development efforts in the early 1980s that found application in early 2011 and continued until now and in the foreseeable future.

Continuous technology scaling in FinFET brought self-heating effects to the forefront of reliability challenges, especially since 14nm technology. Since the silicon volume is limited for heat dissipation (Intel's 2029 node will be 1.4nm, with 12 silicon atoms per channel), the channel's temperature increases during conduction. As a consequence, transistors' basic properties will be degraded, such as drain current being decreased and leakage current becoming greater. An excessive temperature will cause accelerated aging, bias temperature instabilities, and hot carrier injection [19]. The self-heating also causes huge effects on the capacitance of FinFETs. Therefore, technological advancements in manufacturing will result in a shortened lifetime for circuits. The fact that we have modern manufacturing technologies but designs that will never meet the objectives is an absurd situation. This is the exact opposite of the current situation, whereby designs are awaiting proper implementation from manufacturing.

4. Future alternatives due to nanotechnology

In the past, electron conduction in silicon has been the basis of electronics, but since 2000 the gate electrodes of MOSFETs entered the sub-100nm regime, the electronics have entered the nanoelectronic era. These devices are being integrated with new materials and the effects of quantum mechanics are being studied. Before the year 2000, as we reduced the MOSFET size, good things happened automatically like chips running faster and consuming less power. The automatic benefit in early 2000s began to fall apart when chip feature sizes began shrinking below 90nm. In such low dimensions, as electrons move faster and faster, in smaller and smaller silicon circuits, the chips began to heat up [20]. This heat is the fundamental problem and is known as Moore's law's heat death, which requires immediate action. This problem can be resolved in two steps: 1) Manufacturers stopped increasing clock rates, or limited how fast microprocessors execute instructions. A notable fact is that the clock rate hasn't been allowed to increase any further. 2) To meet Moore's performance curve despite clock speed limits, microprocessor chips contained multiple processors, called cores, rather than a single processor. These days, most microprocessors come with dual-core, quad-core, or octa-core capabilities. In fact, four cores with 250MHz speeds may produce the same output as one with 1GHz speeds. The traditional methods of making electronic chips using VLSI technologies are becoming more and more difficult as the feature size limits exceeds the limits of classical physics and requires explanations

based on quantum mechanics. In order to design at such a low scale of integration, high-end equipment is required, which is unaffordable for small VLSI establishments. The advent of nanotechnologies has brought about a transitional period where classical approaches to chip-making are about to give way to emerging methods. Nanotechnology offers the next quantum leap in maximizing the potential of electronic advancements. We will discuss some of the important advances that provide new paradigms for VLSI design in this section.

Nanolithography can be defined as the development of new techniques for making chips, such as etching, writing, printing, *etc.*, of the range of 10^{-6} to 10^{-9} meters. The field can be divided into two main branches: 1) techniques capable of etching molecules, leaving behind the desired structure, and 2) techniques capable of writing the desired structure onto a surface, similar to 3D printing. In the commercially successful fields such as optical, electron beam, nanoimprinting, scanning probe, nanoparticles, nanospheres, proton beams, quantum optics, *etc.* dip pen nanolithography is most promising for use in nanoscale applications in the future [21]. During dip-pen nanolithography, chemical reagents or biological inks are delivered to nanoscale regions of a substrate directly using an atomic-scale tip. This technique is called scanning probe nanopatterning [22]. Today's technologies are so advanced that we need to be able to control matter at atomic levels and build devices with atomic precision. In the scanning tunneling microscope, individual atoms and molecules can be manipulated on surfaces [23], but silicon-based devices require special techniques, such as resist-based strategies, because of covalent bonds. In order to construct any working device such as a transistor with extremely short nanometer length, technologies must be developed that enable an individual atom in a silicon crystal to be placed with atomic precision. An example of such a device is the *"A single-atom transistor"* that was reported in 2012 by Fuechsle *et al.* [24]. They demonstrated a working single atom transistor by combining scanning tunneling microscopy and hydrogen-resist lithography. In this transistor, a phosphorus dopant atom was deterministically placed within a silicon epitaxial device architecture with an accuracy of one lattice site. The transistor was operating at helium temperature and the electron transportation measurements were made at milli-kelvin temperatures. This confirmed that there were discrete quantum energy levels present in the energy spectrum of the phosphorus atom. In 2016, the first report on single-electron tunneling through the oxide tunnel barrier was published at about 0.0037K [25]. Nanoscale cryogenic semiconductor devices are very important for a wide range of applications including quantum computers, since they show faster operation speeds, lower power consumption, smaller leakage currents, lower noise, and lower thermal degradation when compared to their room temperature counterparts [26]. However, the challenge is making devices operate successfully at room temperature.

The charge carriers in the conventional MOSFET are confined to the direction normal to the channel, so that they can move in two dimensions, hence the term 2D carriers. Advances in nanotechnology can enable the creation of designs that allow carriers to move in only one direction through structures known as nanowires. This may be a novel way of designing VLSI *e.g.* MOSFETs based on nanowire fabrications. Nanowire-based field-effect transistors are the most promising way to overcome the limitations of FinFET devices. This design is suitable for gate-all-around architectures that are successful in providing electrostatic control of current with the ultimate reduction of transistor size with low leakage currents [27]. Nanowires' electronic properties, however, are determined by their growth direction, cross-section, and diameter. An interesting fact is that the bandgap of Si nanowire is insensitive to its shape and cross-section. It has been demonstrated [28] that the bandgap occurs at a constant level at different cross-sections under the same surface-to-volume ratio for Si nanowires. This led to a greater likelihood of stability for the devices made with Si-Nano wires. Aviram *et al.* [29] were the first to propose using

molecules as functional electronic devices. The field of molecular electronics has gained much attention since then due to the additional miniaturization strategies it offers. Nanotechnological advancements have allowed electrical measurements to be made on a single molecule. A promising field that utilizes single or multiple arrays of carbon nanotubes instead of silicon nanotubes has been reported to function at room temperature by Trans *et al.* [30]. They built a field-effect transistor consisting of one semiconducting single-wall carbon nanotube connected to two metallic electrodes. This nanotube can be switched from a conducting to an insulating state by applying a voltage to the gate electrode. The device worked well at room temperature, indicating that it may find use in practical applications. The workings of this nanotube transistor can be explained by semi-classical band bending models used for traditional semiconductors. The successful operation of this three-terminal switching device at the molecular level represents a significant development in the field of molecular electronics. Table 1 summarizes the other important models/designs/inventions that could lead to future designs of integrated circuitry.

Table 1: Suggestive studies for futuristic device designs:

Reference(s)	Device Name	Feature length
[31]	Multi-gate SOI MOSFET	5nm to 1.5nm
[32], [33]	Dual material surrounding gate MOSFET	10nm to 1.2nm
[34]	Double step buried oxide SOI MOSFET	10nm (oxide layer)
[35]	Junction less MOSFET with asymmetric gates	13nm to 6nm
[36]	Transparent gate recessed channel MOSFET	30nm to 2nm
[37]	Junction less double gate MOSFET	20nm to 1nm
[38]	Devices based on GaN nanowires	Crossectional 90nm dia.
[39], [40]	Cylindrical surrounding double-gate MOSFET	30nm to 10nm

5. Conclusions

We summarized the VLSI design developments that have enabled IC fabrication to parallel Moore's law in this paper. Since the last decade has seen many challenges in meeting Moore's law demands, this was the decade in focus. VLSI chip design is not as easy in the future since we have tapped into the benefits of traditional silicon MOSFET based methodologies. It is evident from the chronological review presented in this paper that we are in the midst of a transition where traditional technologies will be replaced by new ones based on nanotechnology advances. According to the studies, alternative designs involving nanotechnology will only play a temporal role while far future destiny will be based on biological, molecular, atomic, and spin-based electronic devices, and the term VLSI design will be expunged from chip manufacturing very soon.

References:

[1] G. E. Moore, Proceedings of the IEEE, **86** (1) 82-85 (1998).

[2] M. W. Mitchell, Nature, **530** 145-147 (2016).

[3] B. T. Murthy, D. E. Haggan and W. W. Troutman, Proceedings of the IEEE, **88** (5) 691-703 (2000).

[4] C. Piguet, Low-Power Processor and System on Chip, Boca Raton, (CRC Press, Florida, USA, 2005).

[5] IEEE Andrew S. Grove Award Recipients, https://www.ieee.org/about/awards/bios/grove-recipients.html#2018%20-%20Gurtej%20Singh%20Sandhu

[6] ITRS Reports, International Technology Roadmap for Semiconductors (ITRS), http://www.itrs2.net/itrs-reports.html

[7] C. Demerjian, Intel launches Ivy Bridge amid crushing marketing buzzwords, (2012). https://semiaccurate.com/2012/04/23/intel-launches-ivy-bridge-amid-crushing-marketing-buzzwords/

[8] P. Clarke, Analysis reveal Intel finFETs' triangular structure, EET Asia, (2012). https://archive.eetasia.com/www.eetasia.com/ART_8800667494_499489_NT_02adb57c.HTM

[9] D. Hisamoto, W. C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Brokor and C. Hu, IEEE Transactions on Electron Devices, **47** (12) 2320-2325 (2000).

[10] I. Cutress, Intel's 'Tick-Tock' Seemingly Dead, Becomes 'Process-Architecture-Optimization (2016). https://www.anandtech.com/show/10183/intels-tick-tock-seemingly-dead-becomes-process-architecture-optimization

[11] "Intel Discloses Newest Microarchitecture and 14 Nanometer Manufacturing Process Technical Details," Intel Newsroom (August 11, 2014). https://newsroom.intel.com/news-releases/intel-discloses-newest-microarchitecture-and-14-nanometer-manufacturing-process-technical-details/#gs.bbwwqz

[12] B. Davari, C. Y. Ting, K. Y. Ahn, S. Basavaiah, C. K. Hu, Y. Taur, M. R. Wordeman, O. Aboelfotoh, L. Krusin-Elbaum, R. V. Joshi and M. R. Polcari Symposium on VLSI Technology. Digest of Technical Papers, Karuizawa, Japan, (1987).

[13] B. Yu, L. Chang, S. Ahmed, H. Wang, S. Bell, C.-Y. Yang, C. Tabery, C. Ho, Q. Xiang, T.-J. King, J. Bokor, C. Hu, M.-R. Lin and D. Kyser, Digest. International Electron Devices Meeting, San Francisco, CA, USA, (2002).

[14] "Intel Supports American Innovation with $7 Billion Investment in Next-Generation Semiconductor Factory in Arizona," Intel Newsroom (February 8, 2017). https://newsroom.intel.com/news-releases/intel-supports-american-innovation-7-billion-investment-next-generation-semiconductor-factory-arizona/#gs.c2ogm7

[15] D. I. Cutress, "Intel's Manufacturing Roadmap from 2019 to 2029: Back Porting, 7nm, 5nm, 3nm, 2nm, and 1.4nm," Anand Tech (December 11, 2019). https://www.anandtech.com/show/15217/intels-manufacturing-roadmap-from-2019-to-2029

[16] A. Yasodai and A. V. Ramprasad, International Journal of New Innovations in Engineering and Technology, **14** (1) 67-74, (2020).

[17] International Technology Roadmap for Semiconductors (July 10, 2000) http://www.itrs2.net/

[18] J. Markoff, New York Times (May 4, 2011) https://www.nytimes.com/2011/05/05/science/05chip.html

[19] H. Amrouch, V. M. Santen, O. Prakash, H. Kattan, S. Salamin, S. Thomann and J. Henkel, IEEE 25th International Symposium on On-Line Testing and Robust System Design (IOLTS), Rhodes, Greece, (2019).

[20] S. D. Senturia, Scanning Probe Lithography, New York: Springer Science+Bussiness Media, LLC, (2001).

[21] Z. Durrani, M. Jones, F. Abualnaja, C. Wang, M. Kaestner, S. Lenk, C. Lenk, I. W. Rangelow and A. Andreev, Journal of Applied Physics, vol. **124 (**14) 144502 (2018).

[22] Y. Krivoshapkina, M. Kaestner and I. W. Rangelow, Frontiers of Nanoscience, **11** 497-542 (2016).

[23] D. M. Eigler and E. K. Schweizer, Nature, **344** 524-526 (1990).

[24] M. Fuechsle, J. A. Miwa, S. Mahapatra, H. Ryu, S. Lee, O. Warschkow, L. C. L. Hollenberg, G. Klimeck and M. Y. Simmons, Nature Nanotechnology, **7** 242-246 (2012).

[25] D. I. Bradley, R. E. George, D. Gunnarson, R. P. Haley, H. Heikkinen, Y. A. Pashkin, J. Penttila, J. R. Prance, M. Prunnila, L. Roschier and M. Sarsby, Nature Communications **7** 10455 (2016).

[26] F. A. Mohiyaddin, F. G. Curtis, M. N. Ericson and T. S. Humble, COMSOL Conference, Boston, (2017).

[27] G. Larrieu and X.-L. Han. Nanoscale **5** (6) 2437-2441 (2013).

[28] D. Yao, G. Zhang and B. Li, Nano Letters, **8** (12) 4557-4561 (2008).

[29] A. Aviram and M. A. Ratner, Chemical Physics Letters **29** (2) 277-283(1974).

[30] S. J. Trans, R. M. Verschueren and C. Dekker, Nature **393** 49-52 (1998).

[31] J. P. Colinge, Microelectronic Engineering **84** 2071-2076 (2007).

[32] A. Pal and A. Sarkar, Engineering Science and Technology, an International Journal **17** 205-212 (2014).

[33] F. Djeffal, N. Lakhdar and A. Yousfi, Physica E **44** 339-344 (2011).

[34] A. A. Orouji, S. Heydari and M. Fathipour, Physica E **41** 1665-1668 (2009).

[35] Y. Wang, Y. Tang, L.-l. Sun and F. Cao, Superlattices and Microstructures **97** 8-14 (2016).

[36] A. Kumar, N. Gupta and R. Chaujar, Microelectronics Journal, **49** 36-42 (2016).

[37] Ajay, R. Narang, M. Saxena and M. Gupta, Materials Science in Semiconductor Processing, **71** 240-251 (2017).

[38] C. B. Maliakkal, J. P. Mathew, N. Hatui, A. A. Rahman, M. M. Deshmukh and A. Bhattacharya, Journal of Applied Physics **118** 114301 (2015).

[39] M. A. Uchechukwu and V. M. Srivastva, IEEE Access **8** 121204-121210 (2020).

[40] Y. Pratap and J. H. K. Verma, Silicon, **12** 2619-2626 (2020).

A critical look at the photoelectric effect experiment: a detailed guide for an undergraduate instructor

Ashwin Mohan

Department of Physics, Institute of Chemical Technology,
Matunga, Mumbai 400019, India.
as.mohan@ictmumbai.edu.in

Submitted on 27-10-2020

Abstract

In attempts to explain the photoelectric effect and its theoretical significance, an overview of the experiment using typical sketched current-voltage (I-V) characteristics of phototubes is what is provided in many popular textbooks and lectures that are available on the internet. Here, we shall see that experiments performed on commercially available photoelectric-effect setups used in many teaching laboratories across India show some obvious differences compared to the 'textbook-like' results, and that the interpretation of actual experimental results is complicated by several parameters and physical scenarios that become important from a pedagogical point of view. We shall look at how some issues and caveats related to these experimental findings can be addressed and reasoned by considering the details of the experimental setup used. We shall also see how one can successfully model the general features of typical I-V curves obtained in a photoelectric-effect lab experiment using a fairly quantitative treatment at the undergraduate level.

1 Introduction

Most undergraduate students across India in the physics and engineering streams perform the photoelectric effect (henceforth, PE effect) experiment as part of their laboratory curriculum using commercially available setups. In this experiment, students measure the current-voltage (I-V) characteristics of the phototube for different intensities and frequencies of light. The I-V curve is traced in two different configurations, the forward-bias and the reverse-bias. The reverse-bias characteristic is used to determine the so-called 'stopping potential' for different light frequencies, and the ratio of Plancks constant to the electron charge, 'h/e', a constant,

is thereby found. Also, as part of classroom theory lectures covering modern physics, students are introduced to a schematic experimental setup that demonstrates the PE effect and are taught about the need for invoking the particle nature of light with the concept of 'photons', viewing aspects like intensity and frequency in new perspective. Students are generally shown how the features of typical current versus voltage curves drawn in most textbooks can be rationalized using these newly introduced concepts. Now, an inquiring look at the nature of the I-V curves obtained by students on commercially available PE effect setups, leads to questions and conundrums that are seldom discussed in regular textbooks referred by most students and teachers. Importantly, when closely observed, the I-V curves obtained by students performing the photoelectric experiment using commercially available setups in India differ in some ways from the typical I-V curves sketched in textbooks.

In this article, my aim will be to point out the several issues and caveats that one may encounter while in pursuit of an in-depth understanding of the experimental results of the PE effect experiment on commercially available setups, and direct the reader to useful resources that deal with an advanced level treatment of such issues. I shall qualitatively clarify the experimentally obtained I-V characteristics using commercial PE effect setups in view of aspects like the type of phototube, type of the light source, quality of light filters, geometry of illumination, etc. The article will also include a detailed discussion, at an undergraduate level, required to theoretically understand the PE effect phenomenon so as to provide an expression using which the general features of a typical experimental I-V curve can be simulated. The article will focus on the following points:

- Understand and interpret the current-voltage characteristics of phototubes that are obtained under different circumstances in an actual PE effect experiment using simple concepts from basis solid state physics
- Elucidate to the reader how the interpretation of actual experimental results is complicated by several physical scenarios, thereby illustrating the intricacies one needs to be aware of which are involved while performing this experiment using such commercial setups.
- Enlist and explain assumptions that are relevant for a simplified interpretation of the data.
- Reconcile differences between the simplified curves sketched in textbooks with those obtained experimentally by a student in his or her laboratory session.

2 A brief review

The earliest experimental works on photoelectricity were initiated by Heinrich Hertz in 1887 and Wilhem Hallwachs in 1888, originally known as 'Hallwachs effect' [1, 2].

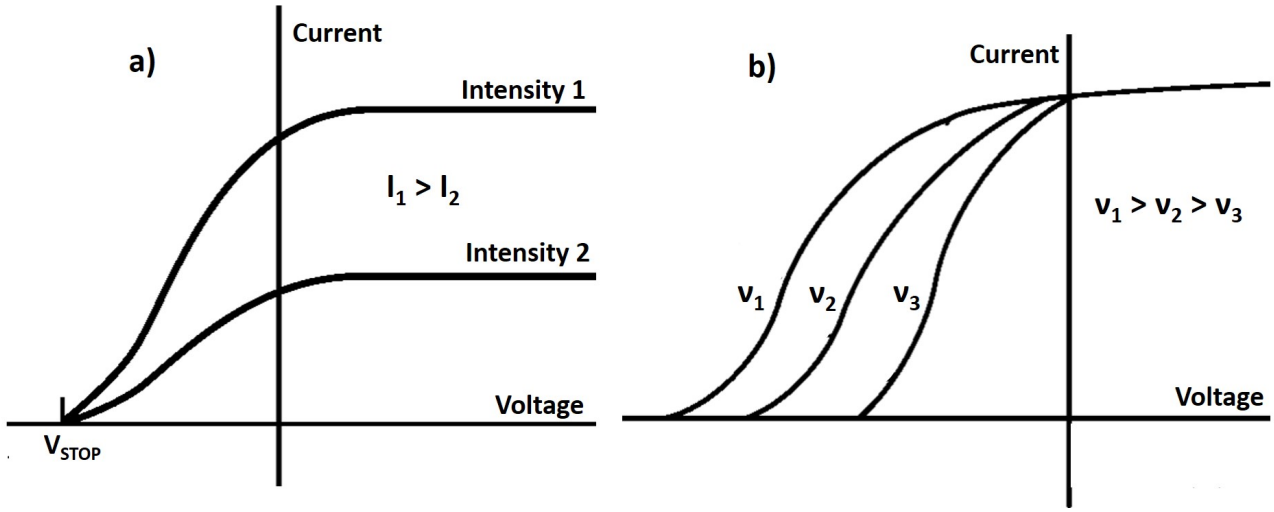


Figure 1: Representative I-V curves in two different configurations, adapted from standard textbooks (a) curves for different light intensities of the same frequency. (b) curves for different light frequencies having the same light intensity.

Phillip Lenard and Robert Millikan also performed series of meticulous experiments that ultimately confirmed Albert Einstein's theoretical hypothesis beyond all doubt [3, 4, 5, 6]. Since then the science of photoelectricity has undergone several developments leading to devices like photodiodes, photo-multiplier tubes, and even photovoltaic cells. Most of us are introduced to photoelectricity through a set of landmark experimental results which were impossible for classical physics to explain, and which led to emergence of the paradigm shifting 'particle-nature-of-light' concept. At the outset, let us briefly review the attempts of six widely used undergraduate-level physics textbooks in explaining the results of the PE effect experiment and associated concepts. These textbooks discuss the results of the experiment by Hertz and Millikan using sketched current versus voltage curves, a typical example of which is shown in Figure 1.

There are two experimental cases that are generally discussed, viz. the 'constant-frequency' case and the 'constant-intensity' case. In the former case I-V curves for different light intensities having the same frequency are given (Figure 1a), and in the latter case I-V curves for different frequencies of a particular light intensity are given (Figure 1b).

In the constant-frequency case, the stopping potential remains same for both intensities, and the saturation current is proportionally higher for the higher intensity. In the constant-intensity case, three different frequencies give the same saturation current but different stopping potentials. The current saturates already in the first quadrant,

as sketched in some books [7, 8, 9]. There is substantial variation in the details of the shape of the curves among textbooks, as these are only sketched curves and as the aim of instruction is not to delve into such details that might be a result of variation in real experimental parameters.

'n_1' photons per unit area per unit time, each having energy $h \cdot \nu_1$ for light of frequency ν_1 are postulated; the intensity, 'I_1' of the light being proportional to $n_1 \cdot h \cdot \nu_1$. This stream of photons is incident on the emitter's surface and these photons transfer their energies to the electrons in the emitter. The energetic photoelectrons emerge from the emitter and reach the collector, thereby constituting a photocurrent. For identical frequencies, the photocurrent is proportional to the incident intensity and the stopping potential is the same for any intensity. No matter how many photoelectrons reach the collector per unit area per unit time, the retarding potential required to stop the most energetic photoelectrons from reaching the collector will remain the same. In the constant-intensity case, the retarding potential required to stop the most energetic photoelectrons from reaching the collector will be different for different incident light frequencies because the photons carry different energies ($h \cdot \nu$) in each case. The forward saturation current remains same regardless of the frequency of incident light. The measured stopping potential increases in proportion to the frequency of the incident photon, and one can thereby determine the value to the Planck's constant. In some books, the gradual increase of photocurrent for increasing accelerating voltages in the forward-bias configuration, like in Figure 1a, is illustrated and explained as follows [10, 11]. Collection of negative 'space-charge' is held responsible for impeding slower photoelectrons from reaching the collector at small accelerating voltages. However, the gradual decline of this 'space-charge' effect with increasing accelerating voltages results in the photocurrent eventually saturating. The above discussion summarizes what a reader can find in typical reference books.

At this point, I would like to flag out that the use of the word 'light-intensity' (I), in almost all reference books, can be misleading, and should be replaced by 'light-flux' (n), a much clearer term. Intensity, in the particle picture, can be varied either by changing the light-flux (n) or by changing the frequency (ν), $I \propto n \cdot \nu$. So, two different frequencies with the same intensity must have different photon-flux and therefore must result in different saturation current. Thus, it is clearer if one talks about constant light-flux for different frequencies, to mean that the number of photons per unit area per unit time remains unchanged.

The following sections will focus on experimental details that are essential to understand the results of the photoelectric effect experiment from a holistic perspective, and attempt at explaining the I-V curves obtained in typical undergraduate experi-

ments using basic concepts of solid state physics that are easily accessible to an undergraduate student. We shall look at obvious differences between the simplified curves sketched in textbooks with those obtained experimentally by a student in his or her laboratory session and attempt to explain them. I shall also point out some issues and caveats that one may encounter while in pursuit of an in-depth understanding of experimental results of the photoelectric effect experiment.

3 The experimental setup

It is necessary for an one to have knowledge of some essential aspects of the experimental setup; it can be invaluable in resolving basic queries related to the experiment, and in rationalizing some features of the I-V curves. Below, we shall talk about the phototube, the collector (anode) and emitter (cathode), light source, and light filters used in typical experimental setups.

3.1 Phototube

Phototubes or Photoelectric cells contain a collector and an emitter, often differently shaped, in a gas-filled or evacuated transparent enclosure. An inert gas filled at low pressure can dramatically alter the sensitivity of the phototube. The first practical devices were invented by Julius Elster and Hans Geitel in 1893. One must keep in mind certain important points regarding the construction and geometry of the phototube. For a detailed account on phototubes the reader may refer to [12, 13].

- There are many possible geometries when in comes to the shape and arrangement of the collector and emitter. Typical geometries that are encountered in commercially available setups are shown in Figure 2. The emitter is a cylindical curved surface, and the collector is a rod or a loop. Other geometries include parallel plate collector and emitter, concentric spherical collector and emitter, etc.
- It is important to know that, for equal magnitude of light fluxes different photosensitive materials show varied response (generate different photocurrents) to radiation of different wavelengths/frequencies. This 'spectral sensitivity' of a phototube can be interpreted as the number of photoelectrons contributing to the photocurrent per quantum of the incident radiation. For example, alkali metals like Li, Na, K show substantial sensitivity below UV frequencies, while a CuO (Copper Oxide) surface is not at all sensitive to any frequency below UV. Most of the photosensitive materials show a gaussian response in a certain frequency range, peaking for a small band of frequencies. The spectral sensitivity for an S4 type phototube is shown in Figure 2 as an example. Such a phototube is highly sensitive in the 'bluish' region of the visible spectrum, and much less sensitive in the 'reddish' region. This is a very important point that will help us understand our experimentally ob-

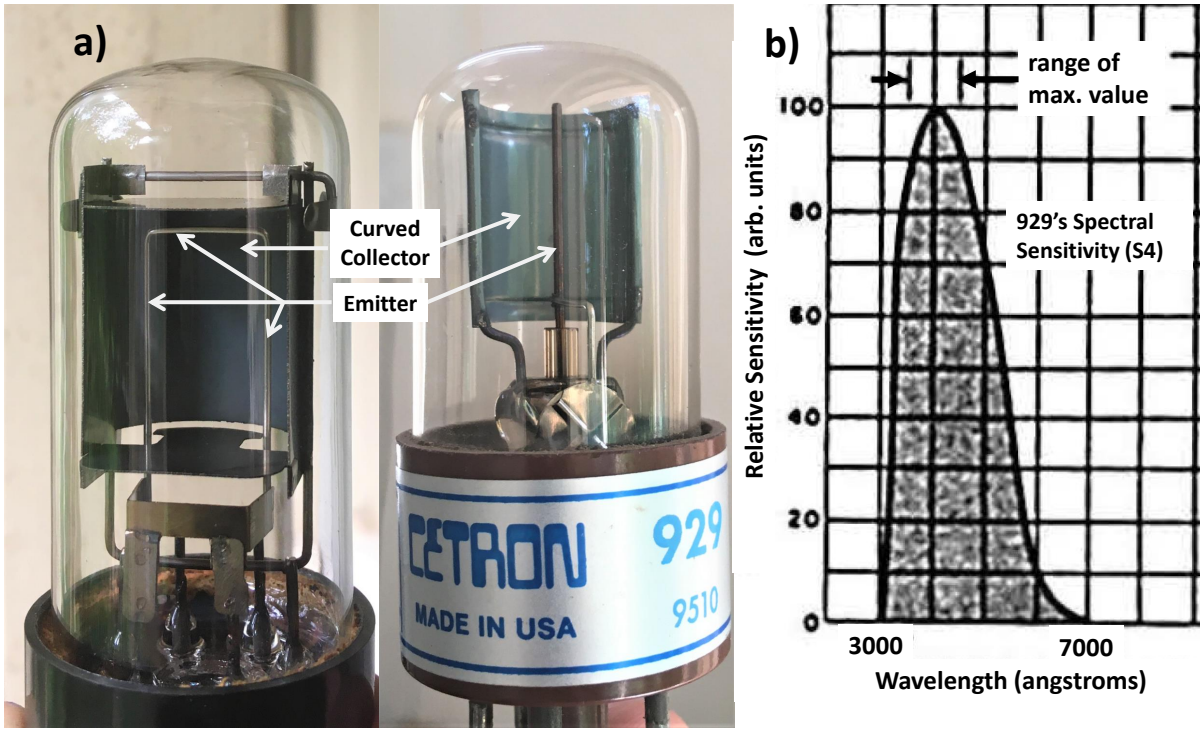


Figure 2: (a) Two different arrangements of the collector and emitter in phototubes used in the two commerically available setups that we have dealt with in this article; Right: Setup 1, Left: Setup 2. (b) The spectral sensitivity of an S4-type phototube implying the dependence of the photocurrent on frequency of incident light; adapted from [12].

tained I-V curves.

• The emission of photoelectrons from a photosensitive surface follows a cosine law, which means that for light incident at any angle on a surface, the number of photoelectrons emitted is proportional to the cosine of the angle between the direction of emission and the normal to the surface [14]. Thus, the maximum number of photoelectrons are emitted perpendicular to the surface. So, a phototube geometry with the anode at the centre of a semi-cylindrical cathode will result in more electrons leaving the cathode going towards the anode than along any other direction, leading to the greatest current for a given voltage and a given light flux on the cathode.

• Given a geometry, a few more factors decide the maximum photocurrent for a phototube. (a) If the size of the collector were larger compared to the thin rod-like structure, many more electrons that earlier whizzed past it without being collected would now contribute to the photocurrent. However, at the same time a large collector would reduce the amount of light that falls on the emitter by shadowing it. (b) Many electrons that miss the anode get accumulated on the insulated enclosure walls of the phototube, which produces an opposing

electric field for the photoelectrons, thereby reducing the photocurrent. This is the 'space charge' effect. (c) The presence of residual gas inside the tube limits the number of electrons reaching the collector owing to scattering events.

• Generally, in phototubes, the collector and emitter are of different material. The collector is generally made of a material that has larger work-function that the emitter. There is a reason for this. Some of the incident light intensity also falls on the collector. If the energy of incident light photon is larger than the work-function of the collector, the collector will emit photoelectrons that will move towards the emitter. This 'reverse-current' will contribute to the net current, especially in the reverse bias condition. The collectors work-function is generally higher than that of the emitter so that the threshold frequency needed for reverse current to flow is high and that for most frequencies there is no reverse current. It is also observed that over time, a residue of the metal that is coated on the emitter builds up on the collector, leading to appreciable reverse current. Thus, in such cases, the zero-current reading in the reverse bias mode will not correspond to no photoelectrons reaching the collector from the emitter, but will indicate the voltage at which the forward and reverse currents cancel out. A correction therefore needs to be taken into account while noting down the stopping potentials in the experiment. This issue, and methods to avoid it has been mentioned from a pedagogic point of view in a few articles that the reader can refer to [15, 16, 17, 18, 19, 20, 21]. Figure 3, adapted from [15], shows a sketched plot of how the net measured photoelectric current can be looked at as a sum of the electron emissions from the emitter and the collector. Many phototubes are designed in a way where incident light is not directly incident on the collector; a 'mask' or a 'lens' is used. Light reflected form the emitter however does reach the collector; this intensity is minimal though. In fact, in some experimental setups like the one used in this article a focussing lens is used to avoid incident light from falling on the collector. The geometry of illumination and its effect on the I-V curves is therefore discussed later in the article.

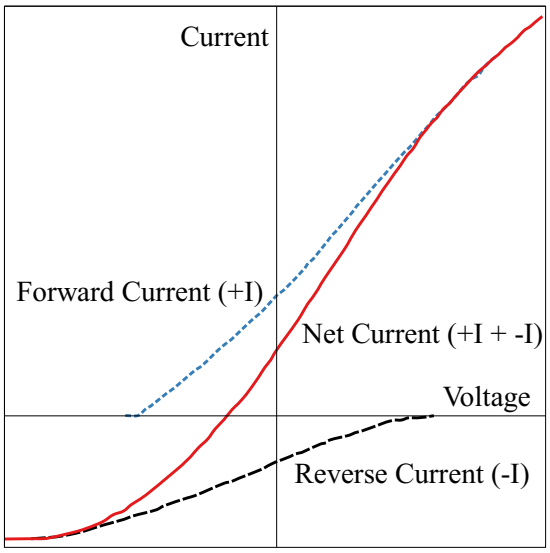


Figure 3: A sketched plot of how the net measured photoelectric current can be looked at as a sum of the electron emissions from the emitter and the collector. Adapted from [15].

• Some fast electrons that reach the collector could be capable of ejecting electrons from the collector material itself and lead to an increase in the photocurrent. These fast electrons, in the case of gas-filled phototubes will be capable of even ionizing the gas molecules and result in a large increase in photocurrent due to an 'avalanche' or 'cascade' of successive ionizations. In this sense, one talks about the 'gas amplification factor for certain types of phototubes where high sensitivity is desired. The physics of such 'secondary emissions become important in devices like photomultiplier tubes.

• Phototubes have a wide range of applications like relay operations in alarm systems, machine safety controls, etc., instruments like photometers, colorimeters, communication systems, industrial control and servomechanisms.

3.2 Light sources

Sources that are typically used in undergraduate labs:

- Tungsten (Sn) halogen lamp
- High pressure Mercury (Hg) vapour lamp
- LED bulbs
- Lasers: semiconductor lasers (cheaper, available in Red, Green, Violet colours) & He-Ne red Laser

The tungsten halogen lamp gives a continuous emission spectrum, the Hg vapour lamp gives a line spectrum; LED bulbs give narrow band monochromatic light output with a Full-width-half-maximum (FWHM) of 20-30 nm; Lasers also give monochromatic output of a much smaller FWHM. In this article we shall discuss results of experiments performed using the four above light sources and point out differences that arise due to variation in the spectral nature of the source.

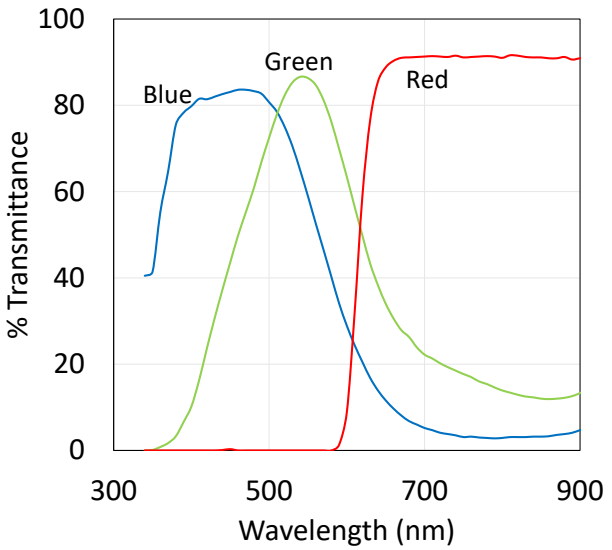


Figure 4: Transmittance curves of filters that are typically supplied with commercially available setups. A specific value of the wavelength on the filter is generally written that is generally used by the student to calculate the Planck's constant in the experiment.

3.3 Light filters

Light filters are provided by the manufacturer of the setup supposedly to isolate small bands of wavelengths to be incident on the emitter. These filters in most setups manufactured in India are made of either

glass or plastic, and must be used with some caution as monochromators in this experiment. Simple measurements using a spectrophotometer give the transmittance curves of these filters as a function of wavelength for tungsten-halogen bulb used as the light source in the spectrophotometer. Figure 4 shows a representative set of transmittance curves that was measured for such filters. Thus, if one uses a source like a tungsten-halogen lamp for performing the photoelectric effect experiment, then one must keep in mind that the light incident on the emitter after passing through such a filter is far from monochromatic; a wide band of around 100 nm is let through. The large uncertainty in wavelength (frequency) makes a reasonable determination of the Plancks constant impossible. Manufacturers often give values of the wavelengths associated with each filter so that one can use these wavelengths to get an acceptable value of '*h*'. It is obvious that they obtain these wavelength values by 'back-calculations' starting from a reasonable value of '*h*', ensuring that the plot of stopping potential versus frequency is a straight line. If such a setup is used, the instructor must be aware of these pitfalls and point them out to the student. The other option is to use a Hg vapour lamp with a line spectrum or LEDs as light sources, in which case a much narrower band of wavelengths is obtained.

4 Understanding the I-V curve

We shall now try to understand the general characteristics of an I-V curve obtained in a typical photoelectric effect experiment using basic concepts of solid state physics and obtain a simple expression that captures its general trend. More detailed attempts at understanding photoemission in an emitter-collector setup can be found in [22, 23, 24]. Let us look closely at the metal surface of the emitter and try to understand the process of photoemission so that a sound explanation of the measured I-V curves becomes possible. Knowledge of some introductory solid-state physics will be required to understand the following section and can be found in standard reference books [25, 26]. The distribution of energy of the electrons residing inside the metal can be simply understood by considering Sommerfeld's free electron model in which all conduction electrons in the interior of the metal experience a constant attractive potential. In this sense, these electrons are not truly "free" (It is important to note that a truly "free" electron cannot absorb a photon completely owing to violation of conservation of energy and momentum.). Earlier to this model, the conduction electrons in a metal was treated as a gas of conduction electrons with Maxwell-Boltzmann distribution of velocities. We can assume that the loosely bound, almost free electrons in the metal (the so called 'conduction electrons') can be considered as a collection of fermions occupying a quasi-continuum of energy levels up till the Fermi energy level

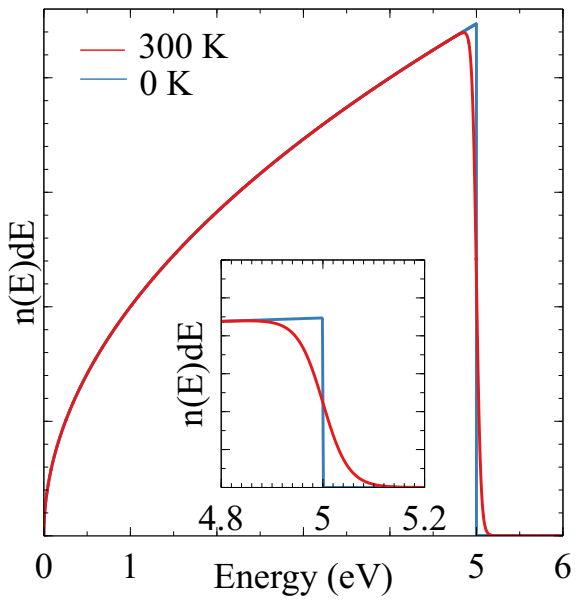


Figure 5: An example plot showing the number of electrons residing in any infinitesimally small energy interval at 0 K and 300 K. Inset: A magnified plot that illustrates the difference in occupation due to temperature.

(E_F). It can be derived from considerations of the free-electron model, how many such quasi-continuous levels or states are available for the electrons to reside in any infinitesimally small energy interval dE. This is called the density of states, $g(E)dE$, and it increases as $\sqrt{E}$. Secondly, the probability of occupancy of these energy levels, i.e. the chance that an electron resides in a particular energy level, in the case of fermions, is given by the Fermi-Dirac statistics. At non-zero temperature the electrons gain some additional energy due to the temperature (for example, $300K \rightarrow 25meV$) and tend to occupy levels above E_F. From here, one can go on to derive an expression for the number of electrons residing in any infinitesimally small energy interval dE, $n(E)dE$. This function is plotted in Figure 5 for $T = 0K$ and $T = 300K$.

The penetration depth of visible light in metals is a few tens of nanometers ($\frac{\lambda}{4\pi k}$); k-extinction co-efficient. So, as far as the photoelectric effect with visible light is concerned, we need to look at the electrons residing only up till a depth of few tens of nanometers of the surface of the metal. If we look at the surface of the metal (tens of nanometers deep) we will see number of electrons with an energy distribution given by $n(E)dE$, instantaneously residing at different depths. Now, let us see what happens when there is an interaction of such electrons with a stream of photons of a certain frequency that is incident on the metal surface. The physics of the interaction of an electron with a photon is a complicated subject to treat theoretically and is not in the scope of a typical undergraduate physics curriculum. Herein, one sees a general result that the nature of interaction is strongly dependent on the energy of the incoming photon. In the case of visible light energy photons, most free electrons will absorb the entire photons energy during the interaction and fully gain the energy $h\nu$ that was possessed by the photon. If such an electron gains enough energy, then it can emerge out of the metal. This is what is termed as 'Photoelectric Emission' (The required energy can also be gained by the electron from the thermal energy due to the temperature

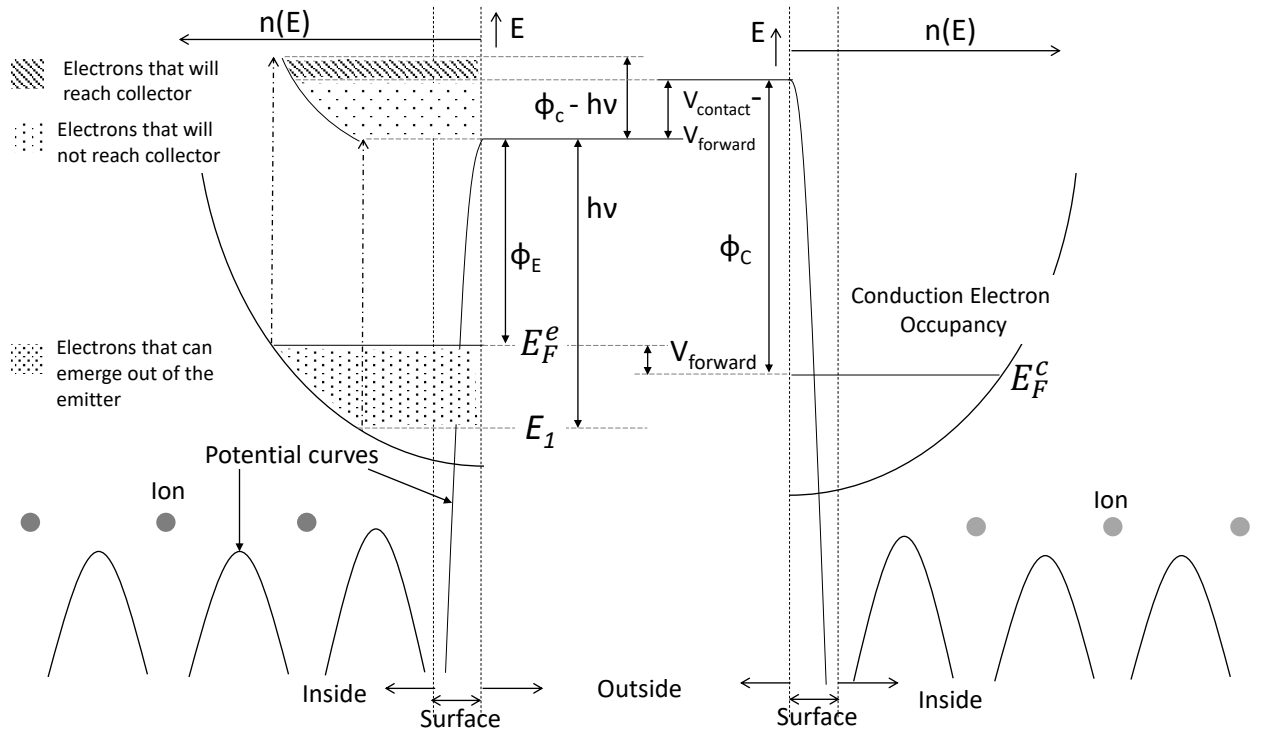


Figure 6: A schematic energy level diagram illustrating the process of photoelectric emission from the emitter to the collector with a small forward-bias voltage. The emitter material is represented on the left and the collector material on the right. Note that the horizontal axes represent the number of electrons when reading the electron occupancy plot, and it represents the depth inside the metal when reading the potential curves. The vertical axis represents energy when reading the electron occupancy plot and the potential curves. The diagram is not to scale.

of the metal; the resulting electrons emerging out of the metal is termed 'Thermionic Emission'). How much energy do the electrons need to emerge out of the metal? Figure 6 depicts an energy level diagram illustrating the process of photoelectric emission. The electrons occupying levels near the fermi-energy are the most energetic electrons inside the metal. These electrons need to gain some more energy, ϕ, to emerge out of the surface till they experience no influence from the metal any more, to the so called 'vacuum'. $\phi_{emitter}$ is termed as the work-function of that metal, and this energy can be gained by an electron by completely absorbing an incident photons energy, i.e. $h\nu$. The excess energy possessed by these electrons after overcoming the barrier at the surface is seen as the kinetic energy of the electrons $h\nu - \phi_{emitter}$. Now, an electron that occupies a level much below the fermi-energy, say E_1, is obviously not so energetic as an electron at the fermi level, and so it requires even more energy to move to the 'vacuum'. One can easily see that this energy that the electron needs to gain should

at least be $(E_F - E_1) + \phi$ for the electron to be able to just emerge out of the metal into 'vacuum'. So, given that we have incident photons each of energy $h\nu$, the lowest energy level in the metal from which an electron can just about emerge out of the metal would be equal to $E_F + \phi - h\nu$; the kinetic energy of such an electron would be zero. And the highest energy level from which an electron emerges out will obviously be E_F. This results in a range of energies which the electrons will have as they emerge out of the metal. So, we will have photoelectrons with a distribution of kinetic energies ranging from a minimum of 'zero' to a maximum of '$h\nu - \phi_{emitter}$'.

When electrons that are situated deep inside the metal encounter an interaction with the photon, they could lose some of their energy owing to collisions with ion cores and other electrons on their way out of the metal. Photoelectrons are ejected from different depths of the emitter surface. Deeper the origin of the photoelectron, higher the probability of it suffering collisions with lattice on its way out, and therefore higher the probability of it having less kinetic energy as it comes out. Photoelectrons from higher up are more likely to have relatively greater kinetic energy; closer to $h\nu - \phi_{emitter}$. This could possibly be another factor that contributes to the distribution of energies of photoelectrons emerging just outside the metal. Such collision processes have low probability of occurrence because these depths (few tens of nanometers) are of the order of the mean free path of conduction electrons in metals; distances that can be traversed by an electron without encountering a collision. The distribution of energies of the electrons owing to the above consideration can therefore be neglected. At this juncture, we must point out that in the above preliminary consideration we have assumed that it is the total energy that the electron possesses which decides whether the electron will emerge out of the metal. However, Nordheim [27] had pointed out that it is only the normal component of the energy that is effective in the escape out of the metal; the normal component of the velocity of the electron that is, and other components do not actually play a role in deciding the distribution of energies of the emerging photoelectrons reaching the collector.

Now, let us see what happens after the photoelectrons exit the metal. Generally, the emitter and collector are made of a different material. When two different metals are connected to each other, electrons flow from on metal to the other till the fermi-energies of the two metals equalize. This process develops a potential difference between the two metals, termed as the Contact Potential (V_C). The above phenomenon is also illustrated in Figure 6 (with the presence of a small forward-bias voltage). As explained before, the collector is generally made of a material that has larger work-function that the emitter. So, when their fermi-energy levels align, the contact potential developed will act as retarding potential for the elec-

trons trying to travel from the emitter towards the collector. All the photoelectrons would experience this retarding potential as they emerge from the emitter. It is evident from Figure 6 that the contact potential is equal to the difference between the work-function of the collector and emitter, i.e. V_C = ϕ_C-ϕ_E. Let us now examine the situation that the photoelectrons experience in different cases. These five cases correspond to five different points/regions on the I-V curve.

• Case 1: If no bias voltage is externally applied between the Emitter and Collector (V_1 = 0): The photoelectrons emerging from the emitter surface, will experience a retarding potential (due to the contact potential V_C) that will make it more difficult for photoelectrons with lower energies to reach the collector. There will nevertheless be some electrons that will have enough energy to overcome this retarding potential and reach the collector, and some current I_1 will be recorded.

• Case 2: If a small forward bias voltage ($V_2^{forward} \ll V_C$) is applied between the Emitter and Collector: The net retarding potential ($V^{retard} = V_C - V_2^{forward}$) experienced by the photoelectrons will be reduced compared to case 1. This will enable some more lower energy electrons to reach the collector and lead to and increase current I_2 being recorded. This way, as the applied forward bias voltage (V) keeps increasing, the retarding potential experienced by the photoelectrons will be lesser and lesser, and the corresponding current recorded (I) keeps increasing.

• Case 3: The forward bias voltage ($V_{saturate}$) applied between the Emitter and Collector equals the Contact potential ($V_{saturate} = V_C$): At this point, the photoelectrons will not experience any net electric field and a current $I_{saturate}$ will be recorded. Most of the electrons will reach the collector, contributing to $I_{saturate}$. It is worth pointing out at this juncture that the photoelectrons that emerge out of the metal and find themselves moving between the Emitter and Collector will encounter occasionally a gas molecule or some boundary inside the phototube. And collision associated loss of energy would reduce the probability of lower energy electrons reaching the collector. The mean free path or the frequency of these collisions will depend on the pressure inside the phototube. If we assume that the pressure is very low inside the phototube, we can say that only few electrons lose their kinetic energy in such collisions, and this phenomenon will have only small effect on the recorded current (see the section on "phototubes" above).

• Case 4: The applied forward bias voltage between the Emitter and the Collector is greater than $V_{saturate}$ of case 3 ($V_{forward} > V_{saturate}$): Beyond $V_{saturate}$ a change of slope is seen in the I-V curve. The recorded current now increases very slowly, behaving almost as the current has saturated. In this region, the photoelectrons only experience an accelerating field that increases with increasing V. Without

any retarding potential, the experimentally recorded current (I) shows only very slow monotonic increase. This slow increase could be attributed to the energy lost by the photoelectrons in collisions made with gas molecules and boundaries of the phototube, as explained in case 3. Practically, a saturation of the photocurrent is measured.

• Case 5: A small reverse bias voltage ($V^{reverse} \ll V_C$) is applied between the Emitter and the Collector: The net retarding potential ($V^{retard} = V^{reverse} + V_C$) will be larger than in case 1. The electrons with relatively lower kinetic energy will find it even more difficult to reach the collector compared to case 1. The recorded current (I) will further decrease. I will keep decreasing as $V^{reverse}$ increases, and eventually go to zero. This voltage ($V^{reverse} = V_{stop}$) will be enough to stop the photoelectrons with maximum energy, i.e. those that were residing at the fermi energy level inside the metal before emerging out. These photoelectrons will have an energy equal to $h\nu - \phi_{emitter}$. We can therefore say $h\nu - \phi_{emitter} = V^{reverse} + V_C$. The applied voltage, $V^{reverse}$, will be less than the actual voltage experienced by the electrons, $V^{reverse}$, by a magnitude of V_C. So, the voltage that is treated as the 'stopping potential' will be more that the applied voltage that is read-off on the voltmeter by V_C.

Looking at cases 1-4, we see qualitatively that the measured photocurrent (I) should increase steadily from zero up till a certain voltage after which it essentially saturates. We now intend to look at a simplified way to model this increase by thinking in lines of the free electron model of metals.

An expression for the total number of conduction electrons present in an energy interval between E and $E + dE$ is given as $n(E)dE = g(E).F(E)dE$, $g(E)$ being the density of states, and $F(E)$ being the Fermi-Dirac occupation function.

$$n(E) = \underbrace{\frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E}}_{\text{g(E)}} \; \underbrace{\frac{1}{exp((E-E_F)/k_BT)+1}}_{\text{F(E)}} \tag{1}$$

where V is the volume of the material, and m is the mass of the electron. This above expression is plotted in Figure 5 for $T = 0K$ and $300K$.

Let us now assume, for instructive purpose, that the effect of the term $F(E)$ is negligible and we can say that the number of electrons goes as $\sqrt{E}$ (this assumption is justified later in the text). Now, for the photoelectric effect we can correlate the incident light energy $h\nu$ with kinetic energy of the photoelectrons, the energy that they possessed inside the metal, and the work-function of the metal emitter (ϕ_E), and can

write the energy E of the electron inside the metal in terms of the kinetic energy (K) it will possess once emerged out of the metal,

$$h\nu = K + \phi_E + (E_F - E) \tag{2}$$

Now, the energy E can be written in terms of the number of electrons n following equation 1 as $E = constant.n^2$. Thus, from equation 2 it follows that,

$$n(K)dK = \sqrt{\frac{K + \phi_E + E_F - h\nu}{C}}dK \tag{3}$$

So, we get an expression for the number of electrons having kinetic energies between K and $K + dK$ ('C' being a constant). As we have explained before, we will have electrons with a distribution of kinetic energies ranging from a minimum of 'zero to a maximum of '$h\nu - \phi_{emitter}$. Therefore, to find the total number of photoelectrons (N) actually reaching the collector for a given potential experienced by these electrons (the actual voltage present between the emitter and collector that is), we need to integrate equation 3. The lower limit of the integration corresponds to those photoelectrons for which the kinetic energy equals the actual net retarding potential ($V_{retard} = V_C - V^{forward}$) that they experience between the collector and emitter (This is because even electrons with non-zero, but very low kinetic energy will be stopped by the retarding potential from reaching the collector and from contributing to the photocurrent). In terms of electron energies this will be '$e(V_C - V^{forward})$; e being the electronic charge. The upper limit will be just '$h\nu - \phi_{emitter}$. So,

$$N = \int_{eV_C - eV^{forward}}^{h\nu - \phi_E} \sqrt{\frac{K + \phi_E + E_F - h\nu}{C}}dK \tag{4}$$

Solving this simple integral, and assuming that the photocurrent measured in the experiment is proportional to N (the number of photoelectrons reaching the collector) for a given potential V ($V^{forward}$ or $V^{reverse}$) experienced by these electrons, we get,

$$I = \frac{2B}{3\sqrt{C}}\left[E_F^{3/2} - ((V_C + \phi_E + E_F - h\nu) - V)^{3/2}\right] \tag{5}$$

where B is a constant ($I = B.N$), and $V \leq V_C$. For voltages above V_C, the current saturates to a constant value equal to,

$$I_{sat} = \frac{2B}{3\sqrt{C}}\left[E_F^{3/2} - (\phi_E + E_F - h\nu)^{3/2}\right] \tag{6}$$

A curve that combines equation 5 and 6 for the two different ranges of V is shown in Figure 7. Equation 5 and 6 will accordingly change if we assume that the anode and cathode are of the same material. V_C will be zero in this case, and the the I-V curve will saturate in the first quadrant as shown in Figure 7, otherwise having similar nature to the previous case. We have hereby captured the general features of the I-V curve using a simple model. Here, we see that the I-V curve for V_C=0 matches the ones sketched in some textbooks, where the curve saturates as it enters the first quadrant. But, we also see how a finite contact potential shifts the curve to the right, thereby changing the stopping potential. Now, we know from case 5 that $h\nu - \phi_{emitter} = V^{stop} + V_C$ when the most energetic photoelectrons are stopped from reaching the collector. If we include $V_C = \phi_C - \phi_E$ in the above expression, we get,

$$h\nu - \phi_C = V^{stop} \tag{7}$$

This means that the work-function that one obtains by looking at the negative y-intercept of the straight line graph of V^{stop} versus ν is the work-function of the collector and not the emitter, as is mostly understood. This misinterpretation has been earlier mentioned in a few pedagogial articles [28, 29, 30, 31, 32, 33], but is overlooked by most. In fact, the above point has been clearly explained even in Robert Millikan's papers [6, 5], but has been neglected in most textbooks. Even the earliest pioneering works by Owen Richardson and Karl

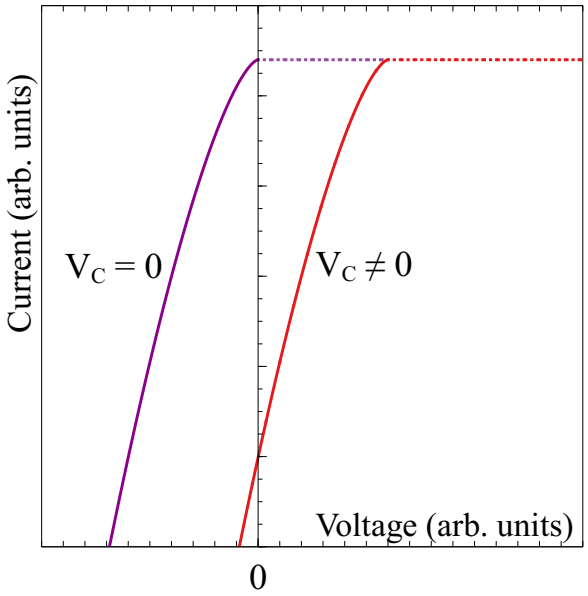


Figure 7: Model I-V curves that combines equation 5 and 6. These curves capture the general trend of experimentally observed I-V curves and helps us understand the effect of the contact potential.

Compton on photoelectric effect took into due account the corrections for the contact potential difference, while calculating maximum retarting potential [34] (Karl Compton was the brother of Arthur Compton who is known for the Compton effect). We must note however, that some important assumptions need to be respected when attempting to understand the I-V curve using a simplistic model as above. Some of these assumptions are:

- The work-function does not change much as a function of the relevant depth inside the metal. In other words, the attractive potential experienced by all the electrons is the same irrespective of

their location in the metal.

- The n Vs E curve is parabolic, and the effect of temperature broadening around fermi energy is neglected, as it will have a very small width of say 0.025 (kT) which is very small compared to the energy $h\nu$ of the photon we are dealing with.
- The probability of photon absorption is equal for the electrons in the metal, independent of their energy in the metal, and independent of the depth of their location in the metal. This might not be completely true.
- The effect of the depth from which electrons are emitted out, is assumed to be negligible on the energy distribution of the electrons.
- The effect of collision of photoelectrons with the gas molecules in the tube, is assumed to be negligible on the energy distribution of the PEs as they travel towards the collector.
- The effect of space charge build up above the emitter is assumed to be negligible on the energy distribution of the photoelectrons as the travel towards the collector.
- All electrons emerge with velocities perpendicular to the surface because the probability that an electron will escape the metal can very well depend on angle of incidence between the electron and the surface.
- The photoelectrons, while travelling from their original position in the metal to the surface after absorbing the photons energy, hardly scatter with other electrons.
- The arrangement and shape of the emitter and collector in the phototube are not considered.
- Surface of emitter is uniform and clean throughout.

5 The reverse current

As previously discussed, generally in phototubes the work-function of the collector is higher than that of the emitter to prevent reverse current from flowing. Sometimes, it is possible that very small amount of the emitters material over time gets deposited on the collector and can cause photoemission from the collector side. This will also happen when the emitter and collector are made of the same material. In the forward biased configuration, the photoelectrons that constitute the reverse current face a retarding potential while moving towards the collector, making it difficult for them to reach the collector, resulting in negligible reverse current. However, when the configuration is reverse biased these photoelectrons experience an accelerating voltage and most of them can reach the emitter. This gives an almost saturated reverse current for reverse voltages (see Figure 3). We have assumed that the behaviour of the reverse cur-

rent curve is similar to that of the forward current curve, much smaller in magnitude than the latter though. In the reverse bias configuration, the forward current (as a result of photoelectrons from emitter reaching the collector) should gradually decrease and reach zero. The forward current is expected to asymptotically reach zero owing to the finite temperature of the emitter, creating a tail-like feature in the I-V curve close to the stopping voltage. This region of the curve corresponds to the photoelectrons that have the largest kinetic energies, i.e. the one that are liberated from near the fermi level. At finite temperatures, small number of electrons in the solid occupy levels higher than the fermi-energy and result in slight smearing of the distribution of electron energies around the fermi level. This smearing of energy levels is directly reflected in the asymptotic tail in the I-V curve for voltages close to the stopping potential. One must remember that a sensitive ammeter will detect a net current combining the forward and the reverse currents. The insufficient isolation of the photocell also causes a small reverse leakage current to flow that is proportional to the applied voltage and in principle adds up to the reverse current. Figure 3 gives an example of a typically obtained I-V curve, which intersects the voltage axis at a certain negative voltage and continues the trend for negative currents until it finally saturates. Such a curve can be thought of as a resultant of two current curves, the forward and the reverse current curves. Hence, to obtain the correct value of the stopping potential, i.e. the potential for which the forward current goes to zero, one must consider the point where the I-V curve saturates rather than considering the point where it intersects the voltage axis. In some cases, the reverse current is negligible, and the I-V curve is seen to asymptotically reach the voltage axis. Here, one can neglect the asymptotic tail and extrapolate the initial curvature to read-off its intersection on the voltage axis.

6 Experimental results

In this section, results of the PE effect experiment performed on two commercial setups manufactured in India by Control Systems & Devices Pvt. Ltd. (Setup 1) and SES Instruments Pvt. Ltd. (Setup 2), both having S4-type phototubes, will be presented and discussed. The main difference between the two setups is the geometry of the collector as shown in Figure 2. Setup 1 has a rod-like collector, and Setup 2 has a rectangular loop-like collector. The maximum voltage that one can apply in Setup 1 is 20 V, and in Setup 2 is 5 V. A focussing lens was used in Setup 2 to direct the incident light preferentially on the emitter, to reduce reverse current in the reverse-bias mode and enable a reasonable estimation of the stopping potential. A commercially available LDR (Light Dependent Resistor) was used to measure the light intensity falling on the phototube. In order to measure the I-V characteristic of the phototube for different frequencies (colours) at

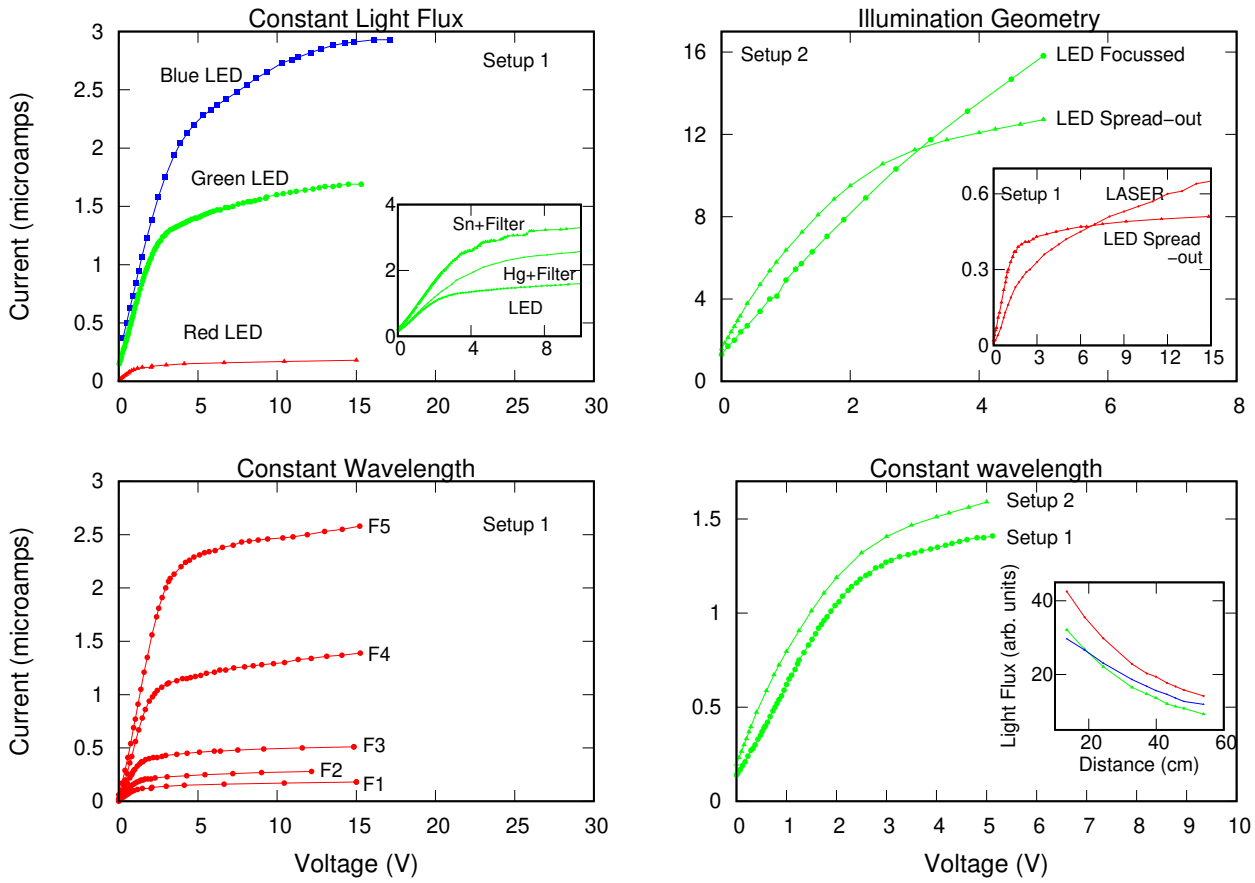


Figure 8: TOP LEFT: The I-V curves are for constant light flux and different LED colours as measured on Setup 1; Inset: I-V curves at constant light flux using different types of light sources in the green region of the visible spectrum. BOTTOM LEFT: All curves are for the same red LED measured at different light fluxes (F1,F2,...,F5) on Setup 1. BOTTOM RIGHT: I-V curves for the same green LED at similar light flux on two different setups; Inset: Relative Light fluxes measured using an LDR as a function of distance from the source for three different LED colours. TOP RIGHT: I-V curves for two different illumination geometries using green LEDs on Setup 1; Inset: I-V curves for two different illumination geometries using Red Laser (focussed case) and Red LED (spread-out case) on Setup 2. See Figure 9 that illustrates different illumination geometries.

the same photon flux, it was ensured that the intensities measured by the LDR at the phototube were in the same ratio as the approximate frequencies of the incoming light ($I \propto n \cdot h\nu$, where n is the photon flux). One should avoid making the mistake of measuring I-V curves at the same intensities as measured by an intensity meter, as this does not correspond to the situation where the light fluxes are same. One desires to measure I-V curves at constant fluxes and not constant intensities in the PE effect experiment.

Figure 8 (Top-Left) shows the I-V curves in forward-bias mode measured using LED

lamps of red, green, and blue colours, keeping the light-flux constant in each case, using Setup 1. We see that even though the light-flux is constant the photocurrent at all voltages increases substantially as the colour (frequency) of the light changes from red to green to blue. Apart from finer details, the observed curves are starkly different from the sketched curves seen in textbooks in terms of saturation currents for different frequencies, and in terms of the region of the first quadrant where the current saturates (see Figure 1). The straightforward explanation for this difference is the non-uniform spectral sensitivity of the phototubes for different wavelengths (see Figure 2). The spectral sensitivity of these S4-type phototubes is far greater in the bluish region than in the reddish region, resulting in observed increase in the photocurrent as a function of colour of incident light. Also, the photocurrent first increases in the first quadrant and saturates beyond a certain voltage, unlike the curves sketched in many textbooks. This is possibly because of the finite contact potential present in the phototubes of these experimental setups, as explained before using a simple theoretical model and as reflected in the simulated I-V curves in Figure 7. There are other experimentally relevant parameters that govern to varying extents the details of the I-V curve like, the range of angles at which the photoelectrons are ejected from the emitter, non-uniformity in the sensitivity of the surface of the emitter, geometry of illumination of incident intensity, configuration and shape of the emitter and collector, etc. The inset of Figure 8 (Top-Left) shows the I-V curves in forward-bias mode measured on Setup 1 using a Tungsten filament lamp with an inexpensive green filter that is generally supplied with all setups, a Mercury vapour lamp and the same green filter, and with a green LED. The incident light-flux was kept same in all three cases. We however see that the photocurrent in the filter+Sn-lamp case is higher at all voltages than that in the case of the filter+Hg-lamp and the green LED. The transmittance for the green and the blue filters used in conjunction with a Sn-lamp are shown in Figure 2, where one sees that the the FWHM for the gaussian-like curve is around 80-120 nm. This means that wavelengths toward the bluish region are also incident on the emitter, thereby increasing the photocurrent for the same light-flux owing to the greater spectral sensitivity of the phototube in the bluish region. The filters happen to be more effective in the case of an Hg-lamp which has a line-like spectral output, and let through less flux in the bluish region. The LED, that is supposed to have a FWHM of 30-40 nm only, gives practically no flux in the bluish region and therefore the least photocurrent. Using commercial setups, one thus has the option of choosing light sources with better and worse monochromaticity.

Figure 8 (Bottom-Left) shows the I-V curves in forward-bias mode using Red LED, for different light fluxes (F1, F2,...,F5) as measured on Setup 1. The photocur-

rent at all voltages is seen to be higher for higher fluxes. In fact, the earliest experiments of Elster and Geitel, Lenard, and Landenburg showed that, as long as there was no change in the spectral quality of the light causing the emission of electrons, the photoelectric current is proportional to the intensity of illumination on the emitting surface. If seen closely, the saturation current for very high fluxes however increases faster than linearly. In fact, it is known that in commercial phototubes often the photocurrent is not proportional to incident light flux at high values due to factors like leaks, charges on the glass walls of the tube and spatial charge.

Figure 8 (Bottom-Right) shows the I-V curves in forward-bias mode using Green LED for the same light-flux, as measured on setup 1 and setup 2 for a comparison between the characteristics of the two phototubes. Apart from a scale factor the features of the curves obtained in the two setups are very similar. One must note that the light-flux decreases as a function of distance from the source in a roughly $1/r^2$ fashion. When the light-flux is measured for different LED colours at a fixed distance it is seen that the magnitude of light-flux depends on the LED colour. So, while measuring the I-V curves of different frequencies one must adjust the distance between the emitter and the source accordingly, rather than keeping the distance fixed for different colours as is instructed in some experimental manuals. This variation with LED colour is shown in the inset of Figure 8 (Bottom-Right).

As mentioned before in this article, many setups are designed so that the incident light is focussed onto the emitter as a spot, without falling on the collector. This is to decrease or avoid the reverse current that makes it difficult to accurately determine the stopping potential. In the experiment, the light flux emitted by the light source is 'focussed' in one case so that only a very small area of the emitter is illuminated by a certain number of photons, as opposed to the 'spread-out' case where roughly the same number of photons as the focussed case are incident on almost all of the emitter's surface area. A simple lens is used in Setup 2 to focus the LED light as shown schematically in Figure 9 (Bottom-Left). The same can also be achieved using the narrow beam of a LASER light. This change in the illumination geometry from spread-out to focussed incidence alters the forward bias I-V characteristics. Figure 8 (Top-Right) shows the I-V curves in forward bias mode using a Green LED in two illumination geometries, spread-out and focussed, on Setup 2. Also, the inset of Figure 8 (Top-Right) shows I-V curves in forward bias mode using Red LED in spread-out illumination geometry and a focussed He-Ne Laser beam, on Setup 1. In both of the above cases, where the light-flux remains the same, we see that, a) at low voltages, forward current for focussed illumination is smaller than that for spread-out illumination, b) at higher voltages, the forward current for focussed illumination is larger

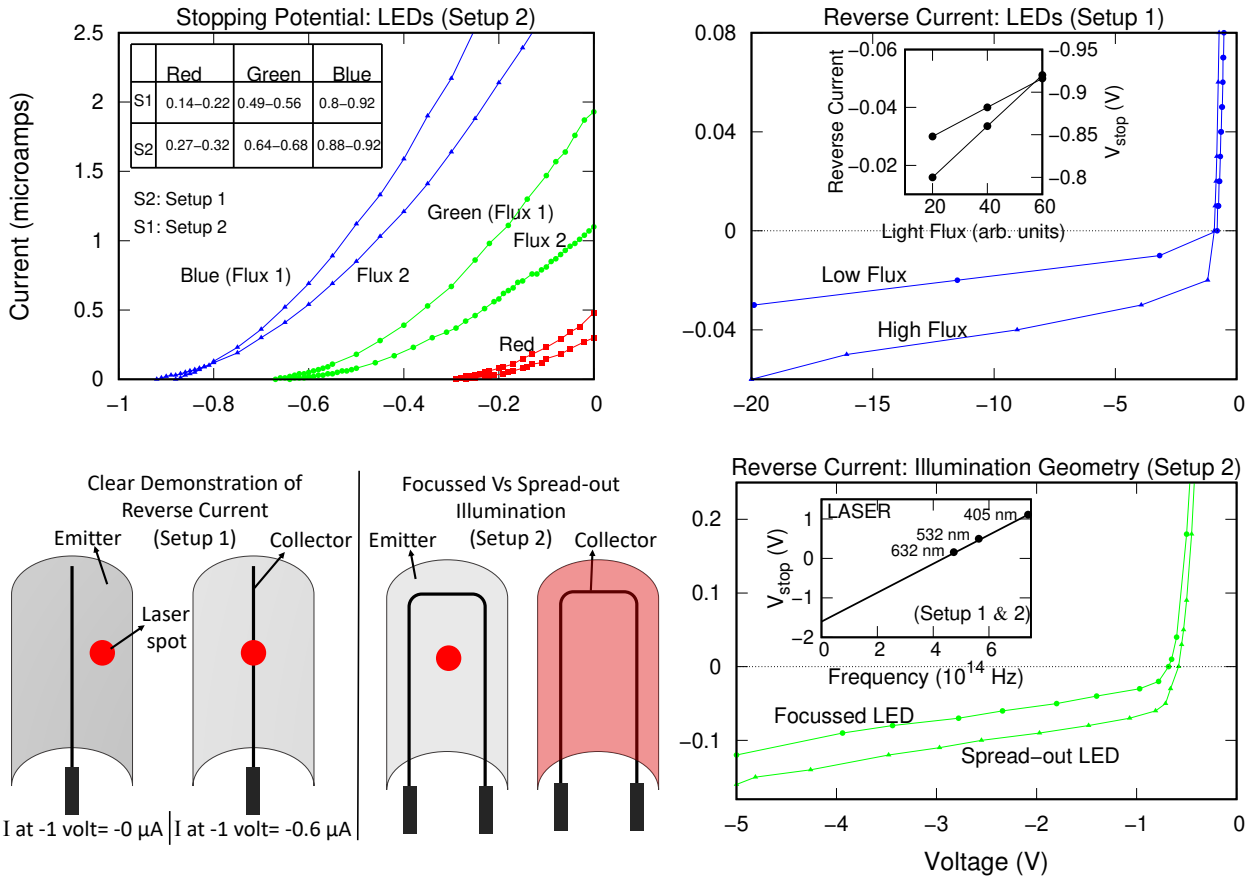


Figure 9: TOP LEFT: I-V curves in the reverse-bias mode for different LED colours at two different light fluxes as measured on Setup 2; Inset Table: The measured range of stopping potentials (in volts) for different light fluxes is shown for three different LEDs as measured on Setup 1 and 2. TOP RIGHT: I-V curves in the reverse-bias mode for two different light fluxes of Blue LED as measured on Setup 1 is shown to illustrate substantial reverse current; Inset: The magnitude of the reverse current and stopping potential is shown as a function of the light flux for the Blue LED case. BOTTOM RIGHT: I-V curves in the reverse-bias mode for two different illumination geometries in the Green LED case is shown to illustrate variation in reverse current and stopping potential with illumination geometry; Inset: Plot of stopping potentials versus frequency of light that shows linear behaviour. This calibration line is identical for both setups when Laser source is used.

than that for spread-out illumination, and does not saturate even at the highest voltage. This means that the extent of illumination on the curved emitter changes the nature of the I-V curve quite substantially. It is therefore advisable to illuminate the entire emitter surface rather than a small spot to be able to see the I-V curve saturating at low and easily attainable forward voltage, and to use the focussed geometry in the reverse-bias condition for reducing the reverse current and the associated error in determina-

tion of the stopping potential.

Figure 9 (Top-Left) shows the I-V curves in reverse-bias for Red, Green and Blue LEDs on Setup 2. We see that for two different light fluxes of the same colour LED, the curves lie very close to each other as they approach the x-axis. This observation is consistent with the theoretical ideas used to explain photoelectric effect. However, if one looks at the x-intercept of these curves a bit more closely one sees that there is a distribution of stopping potentials depending on the light-flux used; higher light-flux giving a higher negative stopping potential. The inset of Figure 9 (Top-Left) enlists the narrow range of stopping potentials measured for LEDs on Setup 1 and Setup 2. So, the stopping potential is seen to be dependent on the light-flux, for a specific LED colour. This was also true in the case of filters+Sn-lamp used as the light source. Such a dependence on the light-flux was however not found in the case of focussed LASERs for different wavelengths; the stopping potentials were constant up to second decimal for different light-fluxes. Therefore, one should keep in mind that the non-monochromatic nature of the incident light for LEDs and filter+Sn-lamp cases, results in a narrow range for stopping potentials. The narrow range of stopping potentials for a specific colour LED/filter therefore leaves an error while choosing a value for the estimation of the Planck's constant. It is therefore advisable to use inexpensive semiconductor LASERs of different wavelengths for an accurate determination of the Planck's constant. This has been done for both the setups, and we get identical stopping potentials for both setups; see inset of Figure 9 (Bottom-Right). Also, LASERs have very well-defined wavelengths compared to LEDs or filters+lamps, further reducing the error when it comes to plotting the x-co-ordinate (frequency) needed in the estimation of the Planck's constant. Here, I have used red (632 nm), green (532 nm), and violet (405 nm) semiconductor LASERs to obtain the stopping potentials in Setup 1 and 2. Figure 9 (Top-Right) shows the I-V curves in reverse-bias mode for Blue LEDs at two different light-fluxes on Setup 1. We see a substantial reverse current, though much smaller than the forward currents, which slowly increases with increasing reverse voltage. As the light-flux increases for a particular LED colour, the reverse current also increases at all negative voltages. Clearly, higher light-flux results in the collector emitting more photoelectrons that travel towards the emitter due to the attractive potential. Also, we see that the stopping potential increases steadily with increasing light-flux (see inset). The net-current curves for different light-fluxes will therefore accordingly cross the x-axis at slightly different voltages depending upon the exact magnitudes of the forward and reverse currents. This results in a range of stopping potentials for light of a particular colour as seen above, and leads to an error in the estimation of the Planck's constant. This

reverse photocurrent will reduce if focussed illumination is used, say Laser, to preferentially allow light to be incident only on the emitter; see Figure 9 (Bottom-Left). It is also better to use low light-fluxes for achieving a better estimate of the true stopping potential owing to smaller reverse currents.

Figure 9 (Bottom-Right) shows the I-V curves in reverse-bias mode using green LEDs for focussed and spread-out illumination geometries. We see that the reverse-current is reduced for the focussed case, and the x-intercept of the curve becomes a better estimate of the stopping potential. It is good to note however that a well-focussed LASER beam is much more effective in reducing the reverse-current than a lens-focussed LED light beam. A clear demonstration of the role of the reverse-current can be shown using a LASER beam spot focussed only on the emitter surface and only on the collector rod; see Figure 9 (Bottom-Left). While the reverse current at -1 volt is zero in the prior case, a large reverse current of -0.6 μA is recorded in the latter case.

7 Conclusion

We have seen how a close inspection of the experimental setup for the photoelectric effect and the resultant I-V curves is invaluable when in comes to a correct and in-depth understanding of the experiment, beyond what is typically mentioned in popular textbooks and many lectures available on the internet. The actual differences between sketched curves found in textbooks and experimentally obtained curves in the lab brings to light the importance of the many experimental parameters that play a role in the photoelectric emission process and its measurement, and possible errors that interpretations of measured data are susceptible to. We have also seen how many of the simple or unexpected experimental findings can be reasoned by looking at the details of the experimental setup used.

Acknowledgments

I acknowledge the enthusiasm of the many students whose curious questioning provoked me to write this article. I would like to thank H. S. Mani, T. R. Govindarajan, Mohan Narayan, and Ayan Banerjee for very enjoyable discussions on this topic. Gratitude is also expressed for the help of Gayatri Thik and Sarika Patil in measurement of transmittance curves.

References

[1] H. Hertz. Ueber einen einfluss des ultravioletten lichtes auf die electrische entladung. *Annalen der Physik*, 267(8):983-1000, 1887.

[2] Wilhelm Hallwachs. Ueber den einfluss des lichtes auf electrostatisch geladene körper. *Annalen der Physik*, 269(2):301-312, 1888.

[3] P. Lenard. Ueber die lichtelektrische wirkung. *Ann. Phys.*, 8:149-198, 1902.

[4] A. Einstein. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322(6):132-148, 1905.

[5] R. A. Millikan. A direct photoelectric determination of planck's "h". *Phys. Rev.*, 7:355-388, 1916.

[6] R. A. Millikan. Einstein's photoelectric equation and contact electromotive force. *Phys. Rev.*, 7:18-32, 1916.

[7] Aurthur Beiser. *Concepts Of Modern Physics*. McGraw-Hill Education (India), 2002.

[8] H. D. Young and R.A. Freedman. *University Physics with Modern Physics*. Pearson Education, 2019.

[9] F. K. Richtmyer and E. H. Kennard. *Introduction to Modern Physics, by F.K. Richtmyer and E.H. Kennard*. International series in physics. McGraw-Hill, 1947.

[10] Robert Resnick and Robert Eisberg. *Quantum Physics: of Atoms, Molecules, Solids, Nuclei and Particles*. Wiley India Pvt. Limited, 2006.

[11] H. C. Verma. *Concepts of Physics*. Bharati Bhawan Publishers, 2019.

[12] A. Schure. *Phototubes*. A Rider publication, no. 166-33. Electronic technology series. J. F. Rider, 1959.

[13] Yu Lukyanov. Survey of vacuum phototubes. *Journal of Technical Physics*, 13(1-2), 1963.

[14] Herbert. E. Ives, A. R. Olpin, and A. L. Johnsrud. The distribution in direction of photoelectrons from alkali metal surfaces. *Phys. Rev.*, 32:57-80, 1928.

[15] Evan. S. Snyder. A correction to the photoelectric current in the plancks constant experiment. *The Physics Teacher*, 23(2):98-99, 1985.

[16] S. K. Chakarvarti and B. L. Sharma. Determination of planck's constant using the photoelectric effect. *Physics Education*, 23(4):249-251, 1988.

[17] Valeria Indelicato, PaolaLa Rocca, Francesco Riggi, Gianluca Santagati, and Gaetano Zappalá. Analysis of LED data for the measurement of planck's constant in the undergraduate laboratory. *European Journal of Physics*, 34(4):819-830, 2013.

[18] R. G Keesing. The measurement of planck's constant using the visible photoelectric effect. *European Journal of Physics*, 2(3):139-149, 1981.

[19] Yaakov Kraftmakher. Photoelectric effect experiment with computer control and data acquisition. *American Journal of Physics*, 74(10):941-943, 2006.

[20] A.C. Melissinos and J. Napolitano *Experiments in Modern Physics*. Experiments in Modern Physics. Elsevier Science, 2003.

[21] Harry. H. Hall and Richard. P. Tuttle. Photoelectric effect and planck's constant in th introductory laboratory. *American Journal of Physics*, 39(1):50-54, 1971.

[22] R. A. Powell. Photoelectric effect: Back to basics. *American Journal of Physics*, 46(10):1046-1051, 1978.

[23] Rosalia Zangara and Elisa Lanzara. Photoemission energy distribution measurements in a simple metal: A modern physics undergraduate laboratory. *American Journal of Physics*, 61(12):1114-1118, 1993.

[24] Lee. A. DuBridge. Theory of the energy distribution of photoelectrons. *Phys. Rev.*, 43:727-741, 1933.

[25] R. J. Singh. *Solid State Physics*. Always Learning. Dorling Kindersley, 2012.

[26] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. HRW international editions. Brooks-Cole Publishing Company, 1976.

[27] Ralph Howard Fowler and L. Nordheim. Electron emission in intense electric fields. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 119(781):173-181, 1928.

[28] Yuval Ben-Abu. Misleading points in the teaching of millikan's experiment on the photoelectric effect. *Physics Education*, 52(4):043006, 2017.

[29] A. N. James. Photoelectric effect, a common fundamental error. *Physics Education*, 8(6):382-384, 1973.

[30] David R. Lloyd. What was measured in Millikan's study of the photoelectric effect? *American Journal of Physics*, 83(9):765-772, 2015.

[31] J. Rudnick and D. S. Tannhauser. Concerning a widespread error in the description of the photoelectric effect. *American Journal of Physics*, 44(8):796-798, 1976.

[32] D. Wong, P. Lee, G. Shenghan, W. Xuezhou, H. Y. Qi and F. S. Kit. The photoelectric effect: experimental confirmation concerning a widespread misconception in the theory. *European Journal of Physics*, 32(4):1059-1064, 2011.

[33] J. Stmad. Die austrittsarbeit beim photoeffekt. *Praxis der Naturwissenschaften-Physik*, 11:343-344, 1980.

[34] O. W. Richardson and Karl. T. Compton. The photoelectric effect. *Science*, 35(907):783-784, 1912.

Design and Implementation of e-CALLISTO Tower Mounted LNA using MMIC

Dnyandev B. Patil[1] and Vijay S. Kale[2]

[1]Research Scholar, Department of Electronics, LVH College, Nashik-03, India.
dnyaneshpatil20@gmail.com

[2]Department of Electronics, KTHM College, Nashik-02,India.
vijaykalesir@rediffmail.com

Submitted on 25-02-2021

Abstract

Low noise amplifiers (LNA) are the building block of communication systems such as Radio, GPS, Mobile phone etc. LNA is the first stage in the most of the receiver system. It amplifies a very low-power signal without degrading its signal to noise ratio (SNR). An LNA with sufficient sensitivity are required to distinguish the residual signal from surrounding noise and interference. LNAs are primarily concerned in front of weak signals receiver and just after antennas. It unites noise figure, reasonable gain, and stability without oscillation over entire useful frequency range. An LNA mounted near antennas is called tower mounted amplifier (TMA). This paper represents the development and installation of TMA using IC CMA-5043+ for e-CALLISO. LNA and Bias Tee are designed using Altium PCB designer. Its simulation results are presented. The results of LNA near to e-CALLISTO receiver and TMA near to LPDA antenna are also presented.

Keywords- e-CALLISTO, LPDA, LNA, TMA. Bias Tee,

1 Introduction

A low noise amplifier (LNA) amplifies a very weak signal without degrading its signal to noise ratio (SNR). Generally, amplifier increases the power of both the signal and the noise present at its input, whereas LNAs are designed to amplify a signal and minimize the noise [1]. In addition, noise can be minimized by using low-noise components, operating point, and circuit topologies. Minimizing additional noise needs to balance power gain and impedance matching [2]. LNAs are primarily concerned in front of weak signals receiver and just after antennas. Low noise amplifiers are the building block of communication systems

and instruments such as Radio, GPS, Mobile phone etc. It is the first stage amplifier in the receiver [3].

A Tower Mounted Amplifier (TMA) is a low-noise amplifier (LNA) which is mounted near antenna in a weather proof and water proof container. Noise figure, gain, non-linearity and impedance matching are the most important parameters in LNA design. According to FriiâĂŹs formula, the first stage of the receiver has main contribution to the overall Noise Figure (NF) of the receiver [4]. The Low Noise Amplifier is similar to class A amplifier. The overall performance of receiver is completely depending on LNA connected in front of it. Main function of LNA is to amplify extremely weak amplitude signals without adding noise, degrading signal, thus maintain the SNR of the system at extremely low power levels [5].

To design TMA, a Monolithic Microwave Integrated Circuit (MMIC) CMA-5043+ is used. This MMIC operates at 45 MHz to 4 GHz microwave frequencies range. MMIC devices typically perform mixing, amplification, and high-frequency switching. Inputs and outputs of this IC are always matched to a 50-ohm characteristic impedance. This makes them easier to use, reducing external matching network when cascading to next stage. MMIC offer several advantages over conventional LNAs, such as reduced size, low cost, and high reproducibility and repeatable performance. All MMICs are designed to operate from a single positive voltage supply [6]. The MMIC and General-Purpose Amplifier (GPA) has powers ranging from 15 to 33 dBm [7]. MMICs are designed with different circuit techniques such as Darlington Pair, Darlington Pair with Active Bias, Discrete with Diode Active Bias, Discrete with integrated current mirror, Field Effect Transistor (FET) operating at self-bias and use three different device technologies such as GaAs Heterostructure Field Effect Transistor (HFET), Indium Gallium Phosphide Heterostructure Bipolar Transistors (InGaP HBT), GaAs Enhancement Mode pseudomorphic High Electron Mobility Transistors (E-pHEMT). [6, 11].

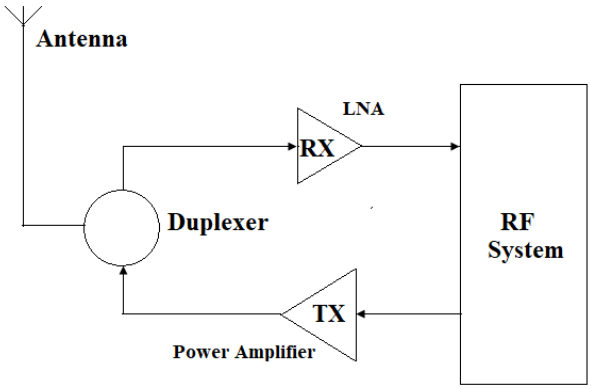


Figure 1: Block diagram of RF system with LNA and Power amplifier.

Figure 1 shows RF transmitter and receiver system with LNA and power amplifier as a building block of the system [9]. The most important function of the LNA is to take the extremely weak signals from the antenna, at the desired frequency range, which is in the order of microvolts and amplify it to a useful voltage level about one-half to one volt. Usually the output is measured in dBm

at 50 Ω impedance system such as 10 uV is -87 dBm and 100 uV equals -67 dBm. [4].

2 Design OF TMA

TMA uses common co-axial cable for RF signal and dc power source. The TMA consist of LNA, optional band pass filter and DC bias tee. Its block diagram is shown in Figure 2. The TMA is connected to the Log-Periodic Dipole Antenna (LPDA) designed for e-CALLISTO receiver.

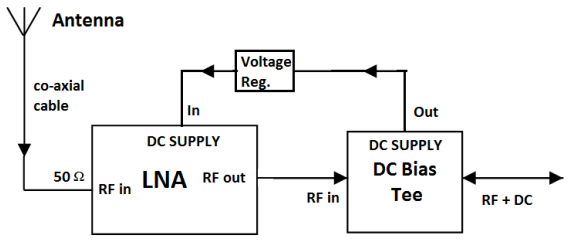


Figure 2: Block diagram of Tower Mounted amplifier (TMA).

LNA: There are many MMIC general purpose amplifier manufacturers are available worldwide such as Analog Devices, NXP, MAXIM, Quorvo, mini-circuits etc. The advantages of such integrated circuits (IC) are low power consumption, high noise immunity, and small size of circuits. The whole LNA is integrated in a small package. Biasing the LNA is most critical part of the design. But most LNA MMICs have an integrated active bias circuit which regulates wide range of supply voltage and frequencies using very few external components [8]. These ICs requires only matching of impedance for desired frequency for optimum gain, linearity and NF. These ICs doesn't require specialized knowledge, skills or equipment for design. In this design we are using CMA-5043+ MMIC. This design operates on +3V supply at only 33 mA current and is internally matched to 50 ohms. The LNA is bonded with a multilayer integrated LTCC substrate and then vacuum sealed under a controlled nitrogen atmosphere with gold-plated covers and eutectic AuSn solder. This MMIC meet the military specification for gross leak, fine leak, thermal shock, vibration, acceleration, mechanical shock, and HTOL [10].

Features of LNA IC CMA-5043+ [10]:

- Ceramic, Hermetically Sealed, Nitrogen filled
- Low profile case, 0.045âĂİ high
- Operates over 0.045 to 4GHz frequency
- Ultra-Low Noise Figure, 0.75 dB
- High Output Power about +21 dBm
- High IP3 and Po at low DC power consumption
- Class 1B HBM ESD rating (500V)
- The working voltage- 3.0v @33 mA and 5v @ 58mA.

All monolithic IC features low external components. Those components are dc blocking capacitors, power supply choke and filter capacitors, shown in Figure 3. Inductor (choke) ADCH 80+ by mini-circuit is used to have stable supply for LNA minimizing dropout of voltage across LNA.

There is no any special design formula for

input and output coupling capacitors.

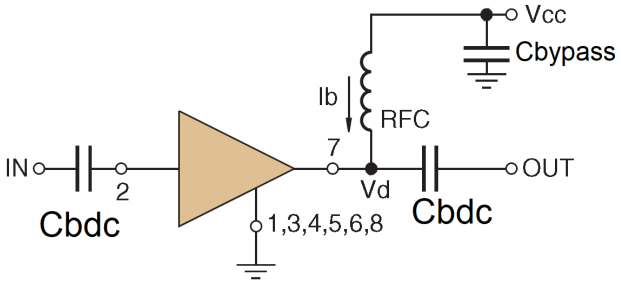


Figure 3: LNA Schematic diagram.

Altium PCB designer:

Protel Systems PVT LTD launched the Printed Circuit Board (PCB) design tool for DOS OS known as Protel PCB. Later it is released by name Altium Designer for windows. The version release in 2007 was the first to offer clearance checking and 3D visualization of PCBs within the software itself. The software features four main functional areas called Schematic designer (capture) , PCB design , Field- Programmable Gate Array (FPGA) development, and data management. Its features also consist of integration with some component distributors which allows searching and order of components and access to manufacturer's data, interactive 3D editing of the board and creating 3D components and export to STEP models for later use. It also provides debugging and simulation of the FPGA board using the VHDL language and checking it for a given input signals. The expected output signals would be generated in Altium Designer [6, 11].

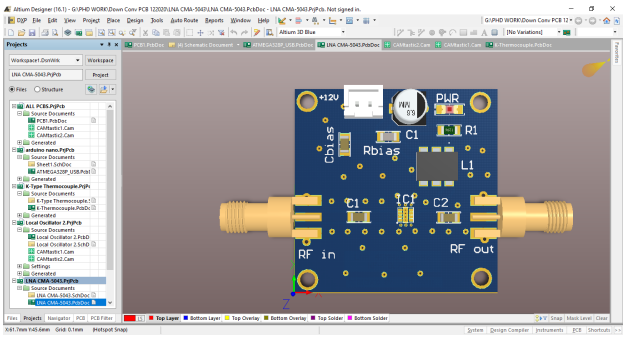


Figure 4: Screenshot of Altium Designer.

Schematic of LNA Design:

LNA PCB is designed in Altium designer software with necessary design of PCB layout given in data-sheet of CMA-5043+ [10].

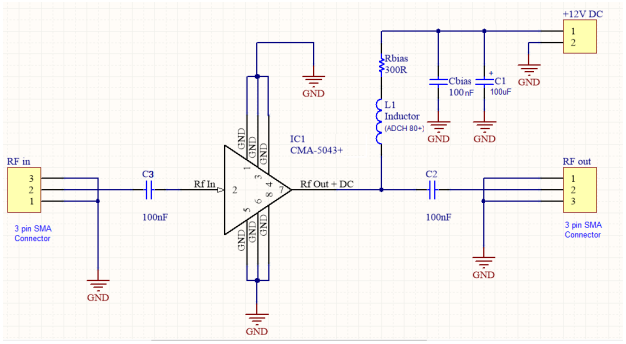


Figure 5: Schematic of low noise amplifier

Capacitor C2 and C3 are used for RF coupling and DC blocking. Also, they are used for impedance matching of input and output. C2 and C3 are chosen such that capacitive reactance, Xc $\ll$ input impedance (50 Ω). Capacitors Cbiase and C1 are used for decoupling and preventing the noise to be induced in circuit. Also,providing pure DC power supply. Resistor R1 is used for current limiting. Its value is chosen by Ohms law such that it will provide 33mA current

when +12v is applied to CMA-5043+. Inductor L1 is for DC supply coupling to internal circuitry and remove the effect of voltage drop across LNA IC. ADCH-80+ is used for this purpose.

Calculations:

Resistance Rbias is selected from the following equation 1.

$$Rbias = \frac{12v - Vcc}{33mA} \quad (1)$$

Design of PCB: As signal will be carried through copper track on pcb, track will act as microstrip coplanar wave guide (mcpw). We need to calculate the width and spacing of track.

Table 1. shows FR-4 material properties

Sr. No	Property	Value
1	Dielectric Constant (ϵr)	3.9 âĂŞ 4.7, 4.4 @ 1 GHz (we take 4.6)
2	Thickness (H) of PCB	1.6mm typical (we use 1.6mm)
3	Copper thickness (T)	0.035 mm typical
4	Loss Tangent (tanD)	0.02 typical

Table 2. Calculations based on FR-4 Material Properties

Sr. No	Property	Value
1	Gap between track and ground plane	0.5mm
2	Track Width	1.89mm
3	Characteristic impedance	50Ω
4	Operating Frequency	6 GHz

Figure 6 and 7 shows the PCB layout and final assembled PCB respectively.

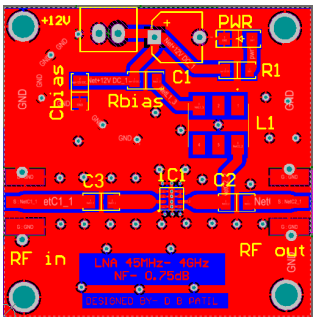


Figure 6: PCB Layout

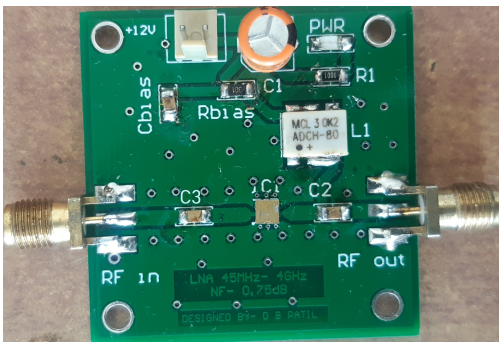


Figure 7: Final Assembled PCB

Dimensions of final PCB (In figure 7)

are 40mm X 40mm.

Test Result at: Supply=+12v, VCC=3v@33mA and R1=300âĎę :

Figure 8 shows the test result of gain versus frequency performance of MMIC CMA-5043+. m1 indicates gain of amplifier at 1GHz frequency which is 18.2dB.

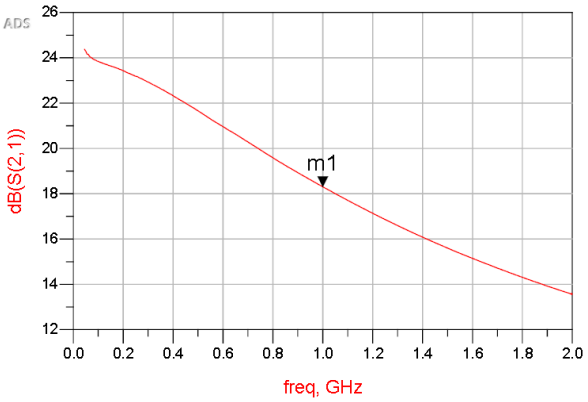


Figure 8: Gain versus frequency graph

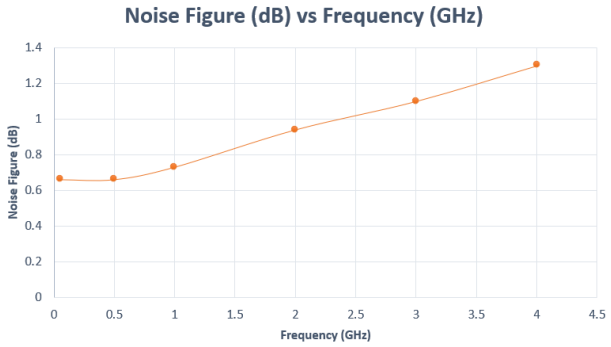


Figure 9: Noise figure of LNA [10]

Figure 9 shows Noise Figure of LNA at frequency range 1 MHz to 6 GHz.

Figure 10 shows measured gain of designed LNA.

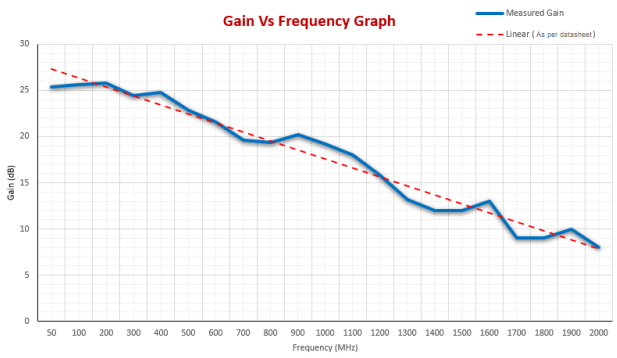


Figure 10: Measured Gain

3 DC Bias TEE

The bias tee is three ports network device, used for dc biasing of some components or devices without disturbing the signal. This dc bias module is used to feed power supply to the LNA and also used to carry DC signal and rf signal from common coaxial cable. It is shown in figure 11.

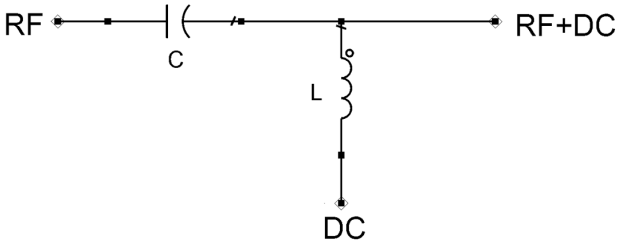


Figure 11: DC Bias Tee

Bias tee are designed for transmission line impedance Z0 of 50 or 75 ohm. Capacitor (XC) chosen such that it is much less than Z0, and the impedance of the inductor (XL) is chosen to be much greater than Z0:

$$X_L = \omega C = 2\pi f L \gg Z_0 \tag{2}$$

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C} \ll Z_0 \tag{3}$$

where ω is the angular frequency. Bias tees are designed to operate over a range of input signal frequencies. The reactance is chosen to have minimal impact at the lowest frequency [12].

For design calculations we will use same parameters described in table 1 and table 2.

Figure 12 and 13 shows the PCB layout of bias tee and complete PCB of bias tee respectively.

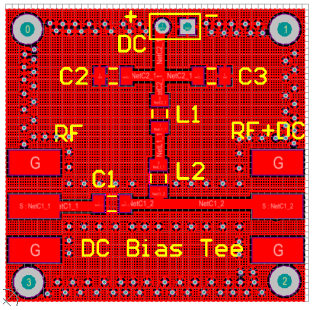


Figure 12: PCB layout of Bias Tee

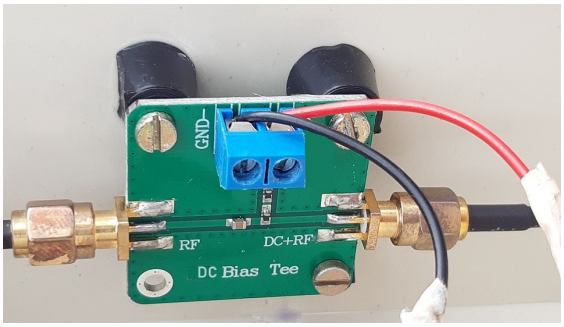


Figure 13: Complete PCB of Bias Tee

Applications of bias tee:

- Biasing amplifiers for remote antenna
- Biasing of laser diodes
- Biasing of active antennas
- DC return
- DC blocking

4 Tests and Results

Figure 14 shows transient analysis test result for input rf signal with dc input for amplifier located at an antenna. We can see the behavior of signal before DC coupling and after DC coupling capacitors for frequency 1MHz which is minimum frequency of input signal. Pink Colour shows Rf + DC signal carrying through cable. Blue colour shows input rf signal of bias tee and green colour shows output rf signal at bias tee near the receiver black colour shows +12v DC supply.

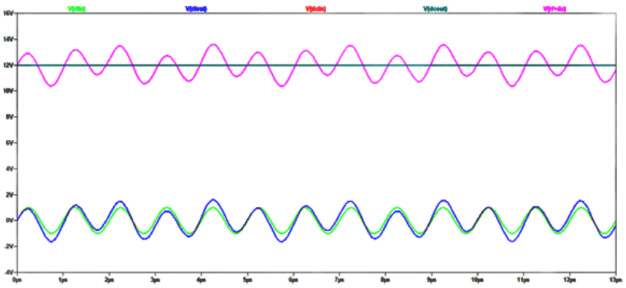


Figure 14: Transient Analysis for RF input at 1MHz

Figure 15 shows the AC analysis from 1MHz to 3GHz. where gain is stable at 20dB after near about 5MHz.

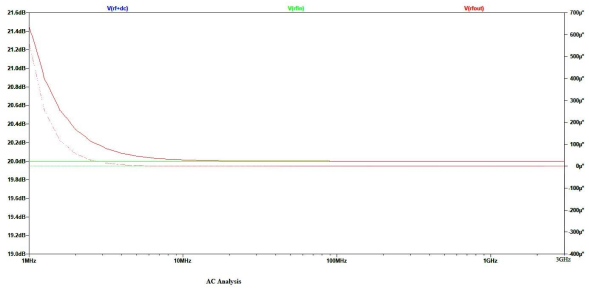


Figure 15: AC Analysis for 1MHz to 3GHz RF input

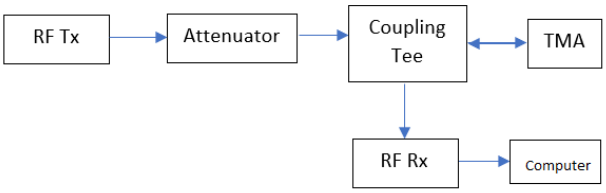


Figure 16: Block diagram for designed TMA SWR setup

System setup is shown in figure 16. Transmitted and reflected signals are captured by rf receiver and then sent to computer, where data is processed and graph is plotted. Figure 17 shows test result of incident waves, reflected wave and VSWR of TMA at low frequency range 50MHz to 1300MHz.

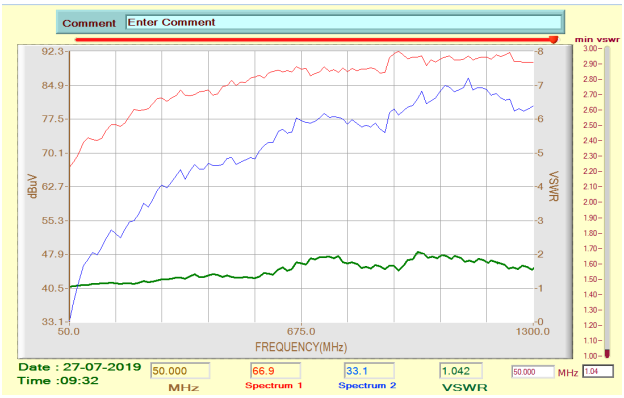


Figure 17: VSWR for designed TMA With Bias Tee upto 1.3Ghz

Designed TMA with Bias Tee

The figure 18 and 19 are the photographs of designed TMA and bias tee respectively. The Figure 20 shows the photograph of installed TMA for e-CALLISTO.

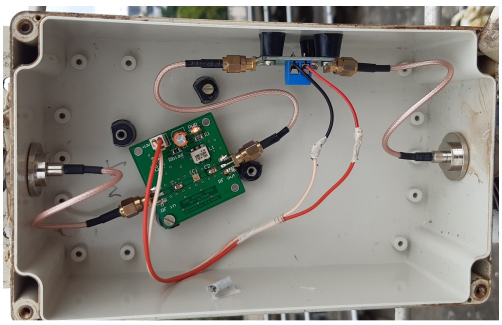


Figure 18: Designed TMA

Figure 19: Designed Bias Tee

Figure 20: Installed TMA.

The experimental setup for radio channel reception is shown in figure 21.

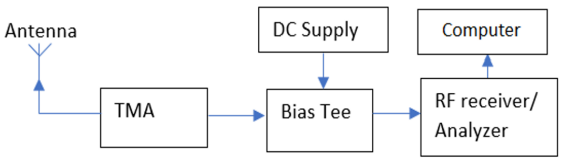


Figure 21: TMA Test setup for radio channel reception

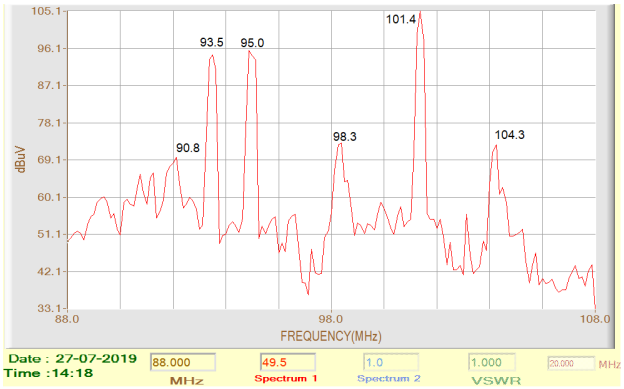


Figure 22: Local FM radio station Signals Captured by TMA.

Figure 22 shows some local FM station signals received by TMA at different frequencies. The TMA was connected to RF receiver.

There are two methods for connecting LNA with antenna. In one, LNA is connected near to receiver while in other LNA is connected near antenna [13, 14]. The major difference is that, for LNA near to receiver uses coaxial cable to receive only signal form antenna and power supply for LNA is provided using separate cable.

Disadvantage in such case is that weak signal is easily affected by noise because signal travels long distance with low amplitude maximising chances of losing information LNA placed near antenna uses same coaxial cable to carry dc supply for LNA and signal form antenna to receiver. Another benefit for using LNA (TMA) near to antenna is reduced noise in received signal. Since, weak signal is amplified near the antenna, the loss of signal is also minimised. Figure 22 and 23 shows comparison between to LNA located near to receiver and antenna respectively. Data is received by e-CALLISTO instrument installed at K.T.H.M. College, Nashik, Maharashtra,India.

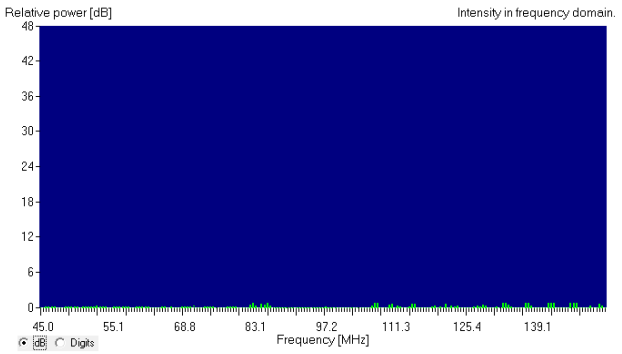


Figure 23: Results of RF power (dB) in frequency domain of LNA located near e-CALLISTO receiver.

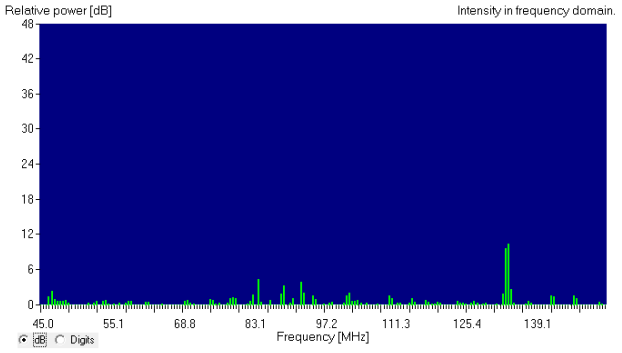


Figure 24: Results of RF power (dB) in frequency domain of e-CALLISTO receiver TMA located near LPDA antenna.

5 Conclusion

The LNA is the building block of communication systems and instruments such as Radio, GPS, Mobile phone etc. It is the first stage amplifier in the most of the receiver systems. It is located in front of receiver. As it deals with the world of unknown signal frequencies, its design is quite complex. The designer must have thorough knowledge of design of LNA and its design rules. But now a days LNAs are manufactured in the Monolithic Microwave Integrated Circuit (MMIC). This MMICs made design very simple such that designer need to just have knowledge of electronics components only. The MMIC requires only few components such as coupling capacitors, AC bias capacitors, inductor and a power supply resistor. The coupling capacitors at input and output of amplifier are used for purpose of transmission line impedance matching and DC blocking only. The PCB design is given in MMICs datasheet. So, the overall design is made very simple. The amplifier uses 50-ohm co-axial cable for carrying both signal and DC supply for LNA. The designed TMA for e-CALLISTO receiver improved the signal quality and strength.

Acknowledgments

This research is supported by the Principal Dr. V. B. Gaikwad and Head, Department of Electronic Science, Dr. V. S. Kale of KTHM College, Nasik, Principal Dr. Digavkar C. G. and Head of Department Dr. Arun Patil, LVH College Panchavati, Nashik, Maharashtra, India. I am thankful of Mr. C. Monstein, Zurich University for his help related to e-CALLISTO and Network. We thank the data center of the e-Callisto network which is hosted by the FHNW Institute for Data Science in Switzerland.

References

[1] D. K. Shaeffer and T. H. Lee A 1.5-V, 1.5-GHz CMOS low noise amplifier. 32 (1957), 745âĂŞ759

[2] J. P. Silver. MOS- Common source LNA-Design Tutorial, www.rfic.uk, 1-11

[3] S. S. Gore and G. M. Phade. Design Challenges and Performance Parameters of Low Noise Amplifier 03 (2013), 204-210

[4] Prameela B and Asha Elizabeth Daniel. Design of Low Noise Amplifier for IEEE standard 802.11b using Cascode and Modified Cascode Techniques. 25 (2016), 443 âĂŞ 449

[5] B. Razavi. RF Microelectronics, second edition ed. Prentice Hall. (2011)

[6] Ting Liu. Design of a Low Noise Amplifier for Wireless Sensor Networks (2011)

[7] Paul Wade. ADL5324 Power Amplifier for Cheap and Simple Microwave Transverters.

[8] https://www.nxp.com/docs/en/application-note/AN3100.pdf

[9] Tim Das Practical Considerations for Low Noise Amplifier Design. (2013).

[10] https://www.minicircuits.com/pdfs/CMA-5043+.pdf.

[11] Moretto Altium releases 3D PCB visualization. (2007), 11-26

[12] James R. Andrews Broadband Coaxial Bias Tee. (2000), 01-03

[13] Yanfen Chen and Fuhong Zhang Design of a 2-Stage Broadband Low Noise Amplifier Based on the Agilent ADS. 719-720, (2015), 862-868

[14] J. P. Silver MOS- Common source LNA-Design Tutorial. 1-11.